

On norms of matrices and some relations between norms and eigenvalues

INAUGURALDISSERTATION

zur Erlangung der philosophischen Doktorwürde, vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

WERNER GAUTSCHI

aus Reinach (Kt. Aargau)

Basel
Buchdruckerei Birkhäuser AG.
1954

On norms of matrices and some relations between norms and eigenvalues

INAUGURALDISSERTATION

zur Erlangung der philosophischen Doktorwürde, vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

WERNER GAUTSCHI

aus Reinach (Kt. Aargau)

Basel
Buchdruckerei Birkhäuser AG.
1954

Genehmigt von der Philosophisch-Naturwissenschaftlichen
Fakultät auf Antrag der

Herren Professoren *A. Ostrowski* und *A. Speiser*.

Basel, den 4. Dezember 1952.

Prof. Dr. *W. Kuhn*

Dekan



1. Teil: Duke Mathematical Journal, vol. 20, pp. 127—140
2. Teil: Duke Mathematical Journal, vol. 20, pp. 375—380
3. Teil: Compositio Mathematica Bd. 11, 2. Heft, 1954, pp. 1—16

THE ASYMPTOTIC BEHAVIOUR OF POWERS OF MATRICES

BY WERNER GAUTSCHI

I. THE MAIN RESULTS

1. **Introduction.** Let $A = (a_{\nu\mu}) (\nu_1\mu = 1, \dots, n)$ be an $n \times n$ matrix with real or complex numbers as elements. By $A^* = A'$ we denote the conjugate-transpose of A and by $\text{tr } A$ the trace $\sum_{\nu=1}^n a_{\nu\nu}$ of A . The *norm* (or *absolute value*) $N(A)$ of A is defined by

$$(1) \quad N(A) = (\text{tr } AA^*)^{\frac{1}{2}} = \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}|^2 \right)^{\frac{1}{2}}.$$

It is well known (J. H. M. Wedderburn [11] or [12; 125]; see also Hardy-Littlewood-Pólya [5; 36]) that

$$(2) \quad N(A + B) \leq N(A) + N(B), \quad N(AB) \leq N(A)N(B), \quad N(\lambda A) = |\lambda| N(A)$$

λ being a scalar and A, B two $n \times n$ matrices.

In (2) of §1 we shall give bounds for the norms of powers $A^p (p = 1, 2, \dots)$. A *lower* bound can readily be found: by a well-known theorem of I. Schur [7] there exists a unitary matrix U which transforms A to a triangular matrix D . The principal diagonal of D consists of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A , not necessarily all distinct, arranged in any desired order. From

$$(3) \quad U^*AU = D, \quad U^*U = UU^* = I$$

it follows that

$$U^*A^pU = D^p, \quad U^*(A^p)^*U = (D^p)^*$$

and

$$\text{tr } A^p(A^p)^* = \text{tr } U^*A^p(A^p)^*U = \text{tr } D^p(D^p)^*.$$

Hence by (1) we have

$$(4) \quad N(A^p) = N(D^p) \geq \left(\sum_{\nu=1}^n |\lambda_{\nu}|^{2p} \right)^{\frac{1}{2}} \quad (p = 1, 2, \dots)$$

with equality for all $p = 1, 2, \dots$, if and only if A is normal, $A^*A = AA^*$.

Suppose now that A can be transformed to the *diagonal form*

$$X^{-1}AX = \Lambda = \text{diag } (\lambda_1, \lambda_2, \dots, \lambda_n)$$

Received February 9, 1952; in revised form, October 18, 1952. This paper is part of the thesis for the doctor of philosophy at the University of Basle, Switzerland.

by a nonsingular matrix X . A necessary and sufficient condition for this is that all elementary divisors of A are simple, *i.e.*, that the minimal polynomial $\phi(\lambda)$ of A has no multiple roots. We then get $\Lambda^p = X^{-1}A^pX$,

$$A^p = X\Lambda^pX^{-1}$$

and, taking norms on both sides and using (2),

$$N(A^p) \leq N(X)N(X^{-1})N(\Lambda^p),$$

that is,

$$(5) \quad N(A^p) \leq c \left(\sum_{\nu=1}^n |\lambda_\nu|^{2p} \right)^{\frac{1}{2}} \quad (p = 1, 2, \dots),$$

where $c > 0$ and only depends on A . (5) can be generalized to cover the case in which $\phi(\lambda)$ has multiple roots. We will show that the right-hand side has to be multiplied by p^{k-1} , where k is, roughly speaking, the highest multiplicity of the roots of $\phi(\lambda)$. The proof is based on a simple application of Sylvester's interpolation formula in the general form.

So far we have only considered *norms* of matrices. We will deal now with the asymptotic behavior of the *individual elements* in the powers of matrices. Necessary and sufficient conditions for the existence of the infinite power $A^\infty = \lim_{p \rightarrow \infty} A^p$ and for A^∞ to be zero have been given by R. Oldenburger [6]. From the point of view of topological algebra O. Taussky [8], O. Taussky and J. Todd [9] have further investigated properties of the class of matrices whose infinite powers are zero. In this paper we shall be concerned in finding upper bounds for the elements of A^p . Matrices B_p which can serve as "majorants" for A^p have been given by Frazer-Duncan-Collar [4; 145-147]. The authors assign to all elements of B_p the same order of magnitude. For certain *triangular* matrices however the order can be graduated as will be shown in 3-4 of §1 by elementary considerations. The result may roughly be characterized as follows: if $D = (d_{\nu\mu})$, $d_{\nu\mu} = 0$ ($\nu > \mu$) is a triangular matrix whose elements along the principal diagonal are arranged in descending order of moduli, the elements in the i -th row of D^p are, if not zero, in modulus at most equal to a constant $\times |d_{ii}|^p$, eventually multiplied by a certain fixed power of p .

In §2 immediate consequences are deduced from Theorem 1 yielding sequences converging towards $\max_\nu |\lambda_\nu|$.

Finally I should like to express my indebtedness to Prof. A. Ostrowski for his valuable criticism and for many helpful suggestions.

2. THEOREM 1. Let A be a (real or complex) $n \times n$ matrix and suppose that not all eigenvalues λ_ν of A are zero. Denote by m_ν the multiplicity of λ_ν in the minimal polynomial of A and put

$$k = \max_{\substack{\nu=1, \dots, n \\ \lambda_\nu \neq 0}} m_\nu.$$

Then we have for a constant $c > 0$ depending only on A

$$(6) \quad 1 \leq \frac{N(A^p)}{\left(\sum_{r=1}^n |\lambda_r|^{2p}\right)^{\frac{1}{2}}} \leq cp^{k-1} \quad (p = 1, 2, \dots).$$

If all λ_r are zero we have $N(A^p) = 0$ ($p \geq l$, $l = m_s$).

Proof. Because of (4) it remains only to prove the right-hand side of (6). If all λ_r are zero, the minimal polynomial of A is λ^l and $A^p = 0$ ($p \geq l$). (By 0 we also denote matrices with zero elements.) We may therefore assume not all λ_r to be zero. Let

$$\phi(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_s)^{m_s}$$

be the minimal polynomial of A with $\lambda_1, \lambda_2, \dots, \lambda_s$ all distinct and put $m = \sum_{\sigma=1}^s m_\sigma$. For any polynomial $f(\lambda)$ we form the $s(m+1) \times m_\sigma$ -matrices Λ_σ and an $(m+1) \times 1$ -matrix Λ_{s+1} :

$$\Lambda_\sigma = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \lambda_\sigma & 1 & \dots & 0 \\ \lambda_\sigma^2 & \frac{2}{1!} \lambda_\sigma & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \lambda_\sigma^{m-1} & \frac{m-1}{1!} \lambda_\sigma^{m-2} & \dots & \cdot \\ f(\lambda_\sigma) & \frac{f'(\lambda_\sigma)}{1!} & \dots & \frac{f^{(m_\sigma-1)}(\lambda_\sigma)}{(m_\sigma-1)!} \end{pmatrix} \quad (\sigma = 1, \dots, s), \quad \Lambda_{s+1} = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \lambda^{m-1} \\ f(\lambda) \end{pmatrix}.$$

The elements in the $(\mu+1)$ -th column of Λ_σ ($\sigma = 1, \dots, s$) are derived from those of the first column by applying the operation $1/\mu!(d/d\lambda_\sigma)^\mu$. Putting together the blocks $\Lambda_1, \Lambda_2, \dots, \Lambda_{s+1}$ into one row we get a square matrix

$$\Lambda = [\Lambda_1, \Lambda_2, \dots, \Lambda_{s+1}].$$

Let us for a moment consider $\det \Lambda$ as a polynomial of the $(s + 1)$ independent variables $\lambda_1, \lambda_2, \dots, \lambda_s, \lambda$; then, as is well known (Turnbull and Aitken [10; 63]), $\det \Lambda$ is divisible by the corresponding "confluent" difference-product Δ ,

$$\begin{aligned} \det \Lambda &\equiv 0 \pmod{\Delta}, & \Delta &= \prod_{\sigma=1}^s (\lambda - \lambda_\sigma)^{m_\sigma} \prod_{\tau>\sigma}^{1,s} (\lambda_\tau - \lambda_\sigma)^{m_\tau m_\sigma} \\ & & &= \phi(\lambda) \prod_{\tau>\sigma}^{1,s} (\lambda_\tau - \lambda_\sigma)^{m_\tau m_\sigma}, \end{aligned}$$

that is,

$$(7) \quad \det \Lambda \equiv 0 \pmod{\phi(\lambda)}.$$

We now expand $\det \Lambda$ in terms of its last row; this gives a relation of the type

$$(8) \quad \det \Lambda = cf(\lambda) + P(\lambda),$$

where

$$c = \prod_{\tau>\sigma}^{1,s} (\lambda_\tau - \lambda_\sigma)^{m_\tau m_\sigma} \neq 0$$

(Turnbull and Aitken [10; 63]; see also Aitken [1; 119–121]) and where $P(\lambda)$ is a polynomial with coefficients in which $f(\lambda_1), f'(\lambda_1), \dots, f(\lambda_2), f'(\lambda_2), \dots$ appear linearly. On the right-hand side of (8) we replace λ by the matrix A ; since $\phi(A) = 0$, it follows from (7) that

$$cf(A) + P(A) = 0.$$

Thus we may write for any polynomial $f(\lambda)$

$$(9) \quad f(A) = \sum_{\sigma=1}^s f_\sigma, \quad f_\sigma = \sum_{\tau=0}^{m_\sigma-1} \frac{f^{(\tau)}(\lambda_\sigma)}{\tau!} C_\tau^{(\sigma)}$$

where $C_\tau^{(\sigma)}$ are matrices *independent of the choice of* $f(\lambda)$ and completely determined by the matrix A . (Compare for this argument Turnbull and Aitken [10; 76–78]. Another proof and more details on formula (9) are given in Wedderburn [12; 27–30].)

We now specialize (9) choosing $f(\lambda) = \lambda^p$ so that

$$(10) \quad A^p = \sum_{\sigma=1}^s f_\sigma, \quad f_\sigma = \sum_{\tau=0}^{m_\sigma-1} \binom{p}{\tau} \lambda_\sigma^{p-\tau} C_\tau^{(\sigma)}.$$

If $\lambda_\sigma \neq 0$ we can write

$$f_\sigma = \lambda_\sigma^p \sum_{\tau=0}^{m_\sigma-1} \binom{p}{\tau} \lambda_\sigma^{-\tau} C_\tau^{(\sigma)},$$

and thus by (2), majorizing the polynomial $\sum \binom{p}{\tau}$ by means of its highest power,

$$(11) \quad N(f_\sigma) \leq c_\sigma |\lambda_\sigma|^p \sum_{\tau=0}^{m_\sigma-1} \binom{p}{\tau} \leq d_\sigma |\lambda_\sigma|^p p^{m_\sigma-1} \quad (p \geq k),$$

c_σ, d_σ being positive constants. Since $f_\sigma = 0$ for $\lambda_\sigma = 0$ and sufficiently large

p , say $p \geq p_0$, it follows from (10), (11) and (2) that for a certain constant $d > 0$

$$(11a) \quad N(A^p) \leq \sum_{\sigma=1}^s N(f_{\sigma}) \leq dp^{k-1} \sum_{\sigma=1}^s |\lambda_{\sigma}|^p \quad (p \geq p_0).$$

Since not all λ_{σ} ($\sigma = 1, \dots, s$) are zero, this upper bound is positive for all integers $p \geq 1$. Hence the constant d can be chosen such that the last inequality also holds for $p = 1, 2, \dots$.

If we change our notations and suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ to be the n (no longer necessarily distinct) eigenvalues of A , we have *a fortiori*

$$N(A^p) \leq dp^{k-1} \sum_{\nu=1}^n |\lambda_{\nu}|^p \quad (p = 1, 2, \dots),$$

which, by Schwarz's inequality, is

$$\leq n^{\frac{1}{2}} dp^{k-1} \left(\sum_{\nu=1}^n |\lambda_{\nu}|^{2p} \right)^{\frac{1}{2}}.$$

This completes the proof of Theorem 1.

We may indicate an alternative, though less elementary proof of (6): adapting the method by which we proved (5) we can transform A to Jordan's canonical form, whose powers are readily calculated. The estimation of their norms can then be carried through similarly as before.

Finally we may remark that the lower bound in (6) can also be attained by nonnormal matrices for certain values of p . Take *e.g.*

$$D = \begin{pmatrix} 1 & 1 \\ 0 & e^{i\psi} \end{pmatrix}$$

and write

$$D^p = \begin{pmatrix} 1 & \epsilon_p \\ 0 & e^{ip\psi} \end{pmatrix};$$

it is easily seen that (compare (16) of the lemma in 4)

$$(12) \quad \epsilon_p = 1 + e^{i\psi} + \dots + e^{i(p-1)\psi} = \frac{e^{ip\psi} - 1}{e^{i\psi} - 1}.$$

Hence if ψ is a rational multiple of 2π , for infinitely many p we have $N(D^p) = 2^{\frac{1}{2}}$ while otherwise $N(D^p)$ may come arbitrarily near to $2^{\frac{1}{2}}$.

On the other hand there exist matrices for which $N(A^p)/(\sum_{\nu=1}^n |\lambda_{\nu}|^{2p})^{\frac{1}{2}}$ is of the same order of magnitude as the upper bound in (6). An example is $D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, where $k = 2$,

$$(13) \quad D^p = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}, \quad N(D^p) = (2 + p^2)^{\frac{1}{2}}.$$

3. Notations and definitions. Let $A = (a_{\nu\mu}) (\nu = 1, \dots, n; \mu = 1, \dots, m)$ be an $n \times m$ matrix with real or complex elements. If $\Delta = \text{diag} (\delta_1, \dots, \delta_n)$ ($\delta_i \neq 0$) is a diagonal matrix of order n , we write A/Δ for $\Delta^{-1}A$. A/Δ is obtained from A by dividing the ν -th row of A by δ_ν ($\nu = 1, \dots, n$). $\langle A \rangle$ will denote the matrix whose elements are $|a_{\nu\mu}|$; if A_1 and A_2 are any $n \times m$ matrices we write $A_1 \ll A_2$, if every element of A_1 is in modulus less than or equal to the corresponding element in A_2 and $\langle A_2 \rangle = A_2$. Clearly $\langle A_1 + A_2 \rangle \ll \langle A_1 \rangle + \langle A_2 \rangle$ and $\langle A_1 A_2 \rangle \ll \langle A_1 \rangle \langle A_2 \rangle$ if both products exist. By I we will denote the unity matrix and by E, E_1 or E_2 (square or rectangular) matrices with elements 1. We call a sequence of $n \times m$ matrices $A_\nu (\nu = 1, 2, \dots)$ *bounded*, if for a certain constant $c > 0$

$$\langle A_\nu \rangle \ll cE \quad (\nu = 1, 2, \dots).$$

A *triangular* matrix is a square matrix with zeros *below* the principal diagonal.

In order to state our second theorem we consider the following seven types of triangular matrices D to each of which we define a " p -th adjoint" diagonal matrix $D^{(p)} (p = 1, 2, \dots)$:

$$\text{I.} \quad D = (d_{\nu\mu}) \quad (\nu, \mu = 1, \dots, m), \quad |d_{11}| > |d_{22}| > \dots > |d_{mm}| > 0.$$

$$D^{(p)} = \text{diag} (d_{11}^p, \dots, d_{mm}^p).$$

$$\text{II.} \quad D = (d_{\nu\mu}) \quad (\nu, \mu = 1, \dots, m), \quad |d_{11}| = |d_{22}| = \dots = |d_{mm}| > 0,$$

$$d_{\nu\nu} \neq d_{\mu\mu} \quad (\nu \neq \mu).$$

$$D^{(p)} = \text{diag} (d_{11}^p, \dots, d_{mm}^p).$$

$$\text{III.} \quad D = (D_{\lambda\kappa}) \quad (\lambda, \kappa = 1, \dots, k),$$

where $D_{\lambda\kappa} (\kappa = 1, \dots, k)$ are triangular matrices of order $n_\kappa > 1$, all diagonal elements of which have the same value d_κ , $D_{\lambda\kappa} (\lambda > \kappa)$ are zero $n_\lambda \times n_\kappa$ matrices and $D_{\lambda\kappa} (\lambda < \kappa)$ arbitrary $n_\lambda \times n_\kappa$ matrices. We further assume

$$|d_1| = |d_2| = \dots = |d_k| > 0, \quad d_\kappa \neq d_\lambda \quad \text{if} \quad \kappa \neq \lambda.$$

In terms of its (scalar) elements we may write $D = (d_{\nu\mu}) (\nu, \mu = 1, \dots, m)$ and define $D^{(p)}$ to be

$$D^{(p)} = \text{diag} (p^{m-1}d_{11}^p, p^{m-2}d_{22}^p, \dots, d_{mm}^p).$$

$$\text{III}'. \quad D = (d_{\nu\mu}) \quad (\nu, \mu = 1, \dots, m), \quad |d_{11}| = \dots = |d_{mm}| > 0.$$

$$D^{(p)} = \text{diag} (p^{m-1}d_{11}^p, p^{m-2}d_{22}^p, \dots, d_{mm}^p).$$

$$\text{IV.} \quad D = \begin{pmatrix} D_1 & B \\ 0 & D_2 \end{pmatrix},$$

where the block D_1 is of type III, the modulus of all diagonal elements being d , $D_2 = (d_{\nu\mu})$ ($\nu, \mu = 1, \dots, m$) is a matrix of type II such that $|d_{\mu\mu}| = d$ ($\mu = 1, \dots, m$) and no $d_{\mu\mu}$ is equal to a diagonal element in D_1 . Then

$$D^{(p)} = \text{diag } (pD_1^{(p)}, D_2^{(p)}).$$

V. $D = (D_{\lambda\kappa})$ ($\lambda, \kappa = 1, \dots, k$),

where the diagonal blocks $D_{\kappa\kappa}$ ($\kappa = 1, \dots, k$) are of one of the types II, III, or IV and d_κ are the moduli of the diagonal elements in $D_{\kappa\kappa}$ such that

$$d_1 > d_2 > \dots > d_k > 0.$$

As in III we suppose $D_{\lambda\kappa} = 0$ ($\lambda > \kappa$) and $D_{\lambda\kappa}$ ($\lambda < \kappa$) to be arbitrary. We define

$$D^{(p)} = \text{diag } (D_{11}^{(p)}, D_{22}^{(p)}, \dots, D_{kk}^{(p)}).$$

VI.
$$D = \begin{pmatrix} D_1 & B \\ 0 & D_2 \end{pmatrix},$$

where D_1 can be of type I-V and $D_2 = (d_{\nu\mu})$ ($\nu, \mu = 1, \dots, m$) is a triangular matrix with zeros along the principal diagonal. Then

$$D^{(p)} = \text{diag } (D_1^{(p)}, 0).$$

4. THEOREM 2. *For a matrix of type I there exists a triangular matrix G with elements 1 along the principal diagonal such that $D^p/D^{(p)} \rightarrow G$ as $p \rightarrow \infty$. For the remaining types II-VI, III' we have*

$$(14) \quad \langle D^p \rangle \ll c \langle D^{(p)} \rangle E \quad (p \geq m),$$

where c is a positive constant, $m = 1$, if D is of type II-V or III', and m is equal to the number of zero diagonal elements, if D is of type VI.

In the cases II-V, III' (14) clearly means that $D^p/D^{(p)}$ ($p = 1, 2, \dots$) is a bounded sequence.

LEMMA. Let $A = \begin{pmatrix} R & T \\ 0 & S \end{pmatrix}$, where R, S are square matrices. Then $A^p = \begin{pmatrix} R^p & T_p \\ 0 & S^p \end{pmatrix}$, where

$$(15) \quad T_p = R^{p-1}T + R^{p-2}TS + \dots + RTS^{p-2} + TS^{p-1},$$

i.e. if R^{-1} exists,

$$(16) \quad T_p = R^{p-1}[T + R^{-1}TS + \dots + (R^{-1})^{p-1}TS^{p-1}].$$

Proof. For $p = 2$ we have, using block-multiplication,

$$T_2 = RT + TS.$$

Hence we may assume our lemma to be true for powers less than p . Then

$$T_p = RT_{p-1} + TS^{p-1} = R[R^{p-2}T + \cdots + TS^{p-2}] + TS^{p-1}$$

which is (15).

Proof of Theorem 2. We first prove the Theorem 2 for matrices of one of the types I, II or III'. Since any matrix of type III is also a matrix of type III' this will include the proof for matrices of type III.

The assertion concerning I, II and III' are clearly true for $m = 1$. We may therefore assume Theorem 2 to be true for matrices of order $< m$ and of type I, II and III'. Let D be of order m ; we divide D in four submatrices, in the following way:

$$D = \begin{pmatrix} d_{11} & \vdots & & B \\ \vdots & \ddots & \dots & \vdots \\ 0 & \vdots & & D_{m-1} \end{pmatrix}.$$

Then we have

$$D^p = \begin{pmatrix} d_{11}^p & \vdots & & B_p \\ \vdots & \ddots & \dots & \vdots \\ 0 & \vdots & & D_{m-1}^p \end{pmatrix}, \quad \frac{D^p}{D^{(p)}} = \begin{pmatrix} d_{11}^p/\delta_{11}^{(p)} & \vdots & & B_p/\delta_{11}^{(p)} \\ \vdots & \ddots & \dots & \vdots \\ 0 & \vdots & & D_{m-1}^p/D_{m-1}^{(p)} \end{pmatrix},$$

where by (16)

$$(17) \quad B_p = d_{11}^{p-1} B \left[I + \frac{D_{m-1}}{d_{11}} + \cdots + \left(\frac{D_{m-1}}{d_{11}} \right)^{p-1} \right],$$

and for $\delta_{11}^{(p)}$ we put d_{11}^p or $p^{m-1}d_{11}^p$ according to the types I, II or III'. It is sufficient to prove the boundedness (or, when D is of type I, the convergence) of $B_p/\delta_{11}^{(p)}$. We consider now two cases:

(a) *Types I, II.* From (17) it follows that

$$(18) \quad \frac{B_p}{d_{11}^p} = B(D_{m-1} - d_{11}I)^{-1} \left[\left(\frac{D_{m-1}}{d_{11}} \right)^p - I \right].$$

For the type I,

$$\left(\frac{D_{m-1}}{d_{11}} \right)^p \rightarrow 0 \quad \text{as} \quad p \rightarrow \infty.$$

(For if all eigenvalues of a matrix A lie inside the unity circle, we have the convergent development $(I - A)^{-1} = \sum_{\nu=0}^{\infty} A^\nu$, so that $A^\nu \rightarrow 0$ ($\nu \rightarrow \infty$).) Therefore

$$(19) \quad \lim_{p \rightarrow \infty} \frac{B_p}{d_{11}^p} = B(d_{11}I - D_{m-1})^{-1},$$

and since, if D is of Type II,

$$\left\langle \frac{D_{m-1}^p}{d_{11}^p} \right\rangle = \left\langle \frac{D_{m-1}^p}{D_{m-1}^{(p)}} \right\rangle \quad (p = 1, 2, \dots),$$

the right-hand side of (18) is bounded by hypothesis.

(b) *Type III'*. Since $\delta_{11}^{(p)} = p^{m-1} d_{11}^p$, it suffices to prove the boundedness of

$$\phi_p = \frac{1}{p^{m-1}} \sum_{\pi=1}^{p-1} \left(\frac{D_{m-1}^{\pi}}{d_{11}^{\pi}} \right)^{\pi} \quad \text{with} \quad p = 1, 2, \dots.$$

By hypothesis for a constant $c > 0$ we have

$$(20) \quad \langle D_{m-1}^{\pi} \rangle \ll c \langle D_{m-1}^{(\pi)} \rangle E, \quad \sum_{\pi=1}^{p-1} \langle D_{m-1}^{\pi} \rangle \ll c \sum_{\pi=1}^{p-1} \langle D_{m-1}^{(\pi)} \rangle E.$$

Put

$$\left\langle \frac{D_{m-1}^{(\pi)}}{d_{11}^{\pi}} \right\rangle = \text{diag} (\pi^{m-2}, \pi^{m-3}, \dots, 1) = \Delta_{m-1}^{(\pi)};$$

then it follows from (20) that

$$\sum_{\pi=1}^{p-1} \left\langle \frac{D_{m-1}^{\pi}}{d_{11}^{\pi}} \right\rangle \ll c \sum_{\pi=1}^{p-1} \Delta_{m-1}^{(\pi)} E \ll cp \Delta_{m-1}^{(p)} E.$$

Hence ϕ_p is bounded.

Type IV. Without loss of generality we may assume $d = 1$; for if $d \neq 1$, write $D = d\bar{D}$; then $D^{(p)} = d^p \bar{D}^{(p)}$, $D^p/D^{(p)} = \bar{D}^p/\bar{D}^{(p)}$.

We put $D^p = \begin{pmatrix} D_1^p & B_p \\ 0 & D_2^p \end{pmatrix}$ and have to show that $B_p/pD_1^{(p)}$ is bounded. By (15) we have

$$B_p = D_1^{p-1}B_1 + D_1^{p-2}B_1D_2 + \dots + B_1D_2^{p-1}.$$

Since D_2^{π} by II is bounded with $\pi = 1, 2, \dots$, we have for a constant $c > 0$

$$\langle B_p \rangle \ll c[\langle D_1^{p-1} \rangle + \dots + \langle D_1 \rangle + I]E.$$

It is therefore sufficient to prove that

$$\frac{1}{pD_1^{(p)}} \sum_{\pi=1}^{p-1} \langle D_1^{\pi} \rangle$$

is bounded, and this follows at once from

$$\langle D_1^{\pi} \rangle \ll c_1 \langle D_1^{(\pi)} \rangle E_1, \quad \sum_{\pi=1}^{p-1} \langle D_1^{\pi} \rangle \ll c_1 \sum_{\pi=1}^{p-1} \langle D_1^{(\pi)} \rangle E_1 \ll c_1 p \langle D_1^{(p)} \rangle E_1,$$

c_1 being a positive constant.

Type V. Our assertion for the type V is now true in the case $k = 1$. For

$k > 1$ we assume it to be true for smaller values of k . Let $D = (D_{\lambda\kappa})$ ($\lambda, \kappa = 1, \dots, k$) and write

$$D = \begin{pmatrix} D_{11} & B \\ 0 & D_{k-1} \end{pmatrix} \quad \frac{D^p}{D^{(p)}} = \begin{pmatrix} D_{11}^p/D_{11}^{(p)} & \dots & B_p/D_{11}^{(p)} \\ \vdots & \ddots & \vdots \\ 0 & \vdots & D_{k-1}^p/D_{k-1}^{(p)} \end{pmatrix}.$$

$$D^p = \begin{pmatrix} D_{11}^p & B_p \\ 0 & D_{k-1}^p \end{pmatrix}$$

By (16) we have

$$\frac{B_p}{D_{11}^{(p)}} = \frac{D_{11}^{p-1}}{D_{11}^{(p)}} [B + D_{11}^{-1} B D_{k-1} + \dots + (D_{11}^{-1})^{p-1} B D_{k-1}^{p-1}].$$

Since $D_{11}^{p-1}/D_{11}^{(p)}$ is bounded we have only to show that

$$\phi_p = \sum_{\pi=1}^{p-1} (D_{11}^{-1})^\pi B D_{k-1}^\pi$$

is bounded. Put $D_{11}^{-1} = 1/d_1 \bar{D}_{11}$; there exists a constant $c_1 > 0$ such that

$$\langle (D_{11}^{-1})^\pi \rangle \ll \frac{c_1}{d_1^\pi} \langle \bar{D}_{11}^{(\pi)} \rangle E_1, \quad \langle D_{k-1}^\pi \rangle \ll c_1 \langle D_{k-1}^{(\pi)} \rangle E_2.$$

The diagonal of $\langle D_{k-1}^{(\pi)} \rangle$ consists of $d_2^\pi, \dots, d_k^\pi$ multiplied by certain powers of π . We enlarge $\langle D_{k-1}^{(\pi)} \rangle$ replacing $d_i^\pi (i = 2, \dots, k)$ by d_2^π and we write $\langle D_{k-1}^{(\pi)} \rangle \ll d_2^\pi \Delta_{k-1}^{(\pi)}$. Hence for a constant $c_2 > 0$ we have

$$\frac{1}{c_2} \langle (D_{11}^{-1})^\pi B D_{k-1}^\pi \rangle \ll \frac{1}{d_1^\pi} \langle \bar{D}_{11}^{(\pi)} \rangle E \langle D_{k-1}^{(\pi)} \rangle E_2 \ll \epsilon^\pi \langle \bar{D}_{11}^{(\pi)} \rangle E \Delta_{k-1}^{(\pi)} E_2,$$

where $0 < \epsilon < 1$, E is rectangular and $\Delta_{k-1}^{(\pi)}$, as well as $\langle \bar{D}_{11}^{(\pi)} \rangle$, is a diagonal matrix containing only powers of π along the diagonal. Thus for a constant $c_3 > 0$ and an integer $s \geq 0$ we can write

$$\epsilon^\pi \langle \bar{D}_{11}^{(\pi)} \rangle E \Delta_{k-1}^{(\pi)} E_2 \ll c_3 \epsilon^\pi \pi^s E,$$

and therefore with $c_4 > 0$

$$\langle \phi_p \rangle \ll c_4 \left[\sum_{\pi=1}^{p-1} \epsilon^\pi \pi^s \right] E.$$

Since the sum in brackets converges as $p \rightarrow \infty$, ϕ_p is bounded with $p = 1, 2, \dots$.

Type VI. Write

$$D^p = \begin{pmatrix} D_1^p & B_p \\ 0 & D_2^p \end{pmatrix};$$

from (16) it follows that

$$(21) \quad \frac{B_p}{D_1^{(p)}} = \frac{D_1^{p-1}}{D_1^{(p)}} [B + D_1^{-1}BD_2 + \cdots + (D_1^{-1})^{p-1}BD_2^{p-1}],$$

and by the Hamilton-Cayley theorem $D_2^p = 0$ ($p \geq m$). Hence the expression in brackets on the right-hand side of (21) is bounded with $p = 1, 2, \cdots$ and for a constant $c > 0$

$$\langle D^p \rangle \ll c \langle D^{(p)} \rangle E \quad (p \geq m).$$

This completes the proof of Theorem 2.

The same example as given at the end of §2 shows that for matrices of type II $D^p/D^{(p)}$ is not convergent in general. In fact, (12) converges only for $\psi \equiv 0 \pmod{2\pi}$. Moreover from (13) it follows that the exponents of p in III cannot be reduced in general.

We should also note that by means of Schur's transformation (3) and (4) an inequality of type (6) could easily be deduced from Theorem 2. The exponent of p however would turn out to be less favorable than in (6).

II. COROLLARIES

1. Put $A^p = (a_{\nu\mu}^{(p)})$ ($\nu, \mu = 1, \cdots, n$),

$$(22) \quad F_p(A) = N^{1/p}(A^p) = [\text{tr } A^p(A^p)^*]^{1/2p} = \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}^{(p)}|^2 \right)^{1/2p} \quad (p = 1, 2, \cdots)$$

and $\omega_A = \max_\nu |\lambda_\nu|$. Then from Theorem 1 we deduce

THEOREM 3. *For every integer $p \geq 1$ we have $\omega_A \leq F_p(A)$ and $F_p(A) \rightarrow \omega_A$ as $p \rightarrow \infty$.*

Proof. We may assume $\omega_A > 0$; then by Theorem 1

$$(23) \quad \omega_A \leq \left(\sum_{\nu=1}^n |\lambda_\nu|^{2p} \right)^{1/2p} \leq F_p(A) \leq (cp^{k-1})^{1/p} \left(\sum_{\nu=1}^n |\lambda_\nu|^{2p} \right)^{1/2p},$$

and our second assertion follows by making $p \rightarrow \infty$.

This theorem has been stated by A. B. Farnell [3] in the case that p runs through powers of 2. For the convergence of $F_p(A)$ he has indicated a proof for matrices of second order, valid also when p runs through all integers. In a way the theorem is analogous to the fact that for an essentially bounded function, integrable in the sense of Lebesgue

$$N_p(f) = \left[\int |f|^p dx \right]^{1/p} \rightarrow \max |f| \quad (p \rightarrow \infty).$$

The analogy however is not perfect, since the numbers $F_p(A)$ cannot be considered as norms for matrices the triangular inequality being not satisfied in general. For from

$$F_p(A + B) \leq F_p(A) + F_p(B) \quad (p = 1, 2, \dots),$$

making $p \rightarrow \infty$, it would follow $\omega_{A+B} \leq \omega_A + \omega_B$ which is false in the case

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad A + B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

where $\omega_A = \omega_B = 0$, $\omega_{A+B} = 1$.

The following corollary comes out of Theorem 3 particularly easily:

COROLLARY. For any square matrix A let $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_n^2$ ($\rho_r \geq 0$) be the eigenvalues of AA^* and put $M(A) = \rho_1$. Then

$$\omega_A \leq M^{1/p}(A^p) \rightarrow \omega_A \quad (p \rightarrow \infty).$$

In fact, we have

$$\omega_A \leq M(A) \leq \left(\sum_{r=1}^n \rho_r^2 \right)^{\frac{1}{2}} = (\text{tr } AA^*)^{\frac{1}{2}},$$

and therefore by (22)

$$\omega_A \leq M^{1/p}(A^p) \leq [\text{tr } A^p(A^p)^*]^{1/2p} = F_p(A).$$

2. As the values $F_p(A)$ are of rather intricate structure we give a sequence $G_p(A)$, equally convergent to ω_A , whose elements are somewhat simpler to calculate. Put

$$R_p = \max_{\nu} \sum_{\mu=1}^n |a_{\nu\mu}^{(p)}|, \quad T_p = \max_{\mu} \sum_{\nu=1}^n |a_{\nu\mu}^{(p)}|,$$

$$G_p(A) = R_p^{1/p} \quad (p = 1, 2, \dots).$$

THEOREM 4. For every integer $p \geq 1$ we have $G_p(A) \geq \omega_A$ and $G_p(A) \rightarrow \omega_A$ as $p \rightarrow \infty$.

Proof. The inequality follows immediately from the well-known inequality $\omega_A \leq R_1$, applied to the matrix A^p . For any matrix $C = (c_{\nu\mu})$ ($\nu, \mu = 1, 2, \dots, n$) we have using Schwarz's inequality

$$(24) \quad \sum_{\mu=1}^n |c_{\nu\mu}| \leq n^{\frac{1}{2}} \left(\sum_{\mu=1}^n |c_{\nu\mu}|^2 \right)^{\frac{1}{2}} \leq n^{\frac{1}{2}} \left(\sum_{\nu, \mu=1}^n |c_{\nu\mu}|^2 \right)^{\frac{1}{2}} = n^{\frac{1}{2}} N(C).$$

We apply (24) to the matrix $C = A^p$ and take ν such that $\sum_{\mu=1}^n |a_{\nu\mu}^{(p)}| = R_p$. This gives

$$\omega_A \leq R_p^{1/p} \leq n^{1/2p} F_p(A)$$

and, by making $p \rightarrow \infty$, our assertion follows from Theorem 3.

Since A and its transpose A' have the same eigenvalues, Theorem 4 holds for the sequence $T_p^{1/p}$ as well; in particular

$$H_p(A) = \min^{1/p} (R_p, T_p)$$

tends to ω_A as $p \rightarrow \infty$.

3. It is natural to ask under which conditions the sequences $F_p(A)$ and $G_p(A)$ will *monotonically* decrease towards ω_A . In this direction we prove

THEOREM 5. *If $p_\nu > 0$ ($\nu = 1, 2, \dots$) is a strictly increasing sequence of integers, such that every p_ν is divisible by $p_{\nu-1}$ ($\nu = 2, 3, \dots$), then $F_{p_\nu}(A)$, $G_{p_\nu}(A)$ are monotonically decreasing towards ω_A as $\nu \rightarrow \infty$. The monotony of $F_{p_\nu}(A)$ and $G_{p_\nu}(A)$ has been proved by A. B. Farnell [3], A. Brauer [2] respectively for the special case $p_\nu = 2^{\nu-1}$.*

Proof. We first prove that for any integer $k \geq 1$

$$(25) \quad F_{k+1}(A) \leq F_1(A) \left[\frac{F_k(A)}{F_1(A)} \right]^{k/(k+1)} \quad (k = 1, 2, \dots).$$

In fact, applying Schwarz's inequality we have

$$|a_{\nu\mu}^{(k+1)}|^2 = \left| \sum_{\lambda=1}^n a_{\nu\lambda}^{(k)} a_{\lambda\mu} \right|^2 \leq \left(\sum_{\lambda=1}^n |a_{\nu\lambda}^{(k)}| |a_{\lambda\mu}| \right)^2 \leq \sum_{\lambda, \kappa=1}^n |a_{\nu\lambda}^{(k)}|^2 |a_{\kappa\mu}|^2.$$

Summing over ν and μ we get

$$\sum_{\nu, \mu=1}^n |a_{\nu\mu}^{(k+1)}|^2 \leq \left(\sum_{\kappa, \mu=1}^n |a_{\kappa\mu}|^2 \right) \left(\sum_{\nu, \lambda=1}^n |a_{\nu\lambda}^{(k)}|^2 \right),$$

that is,

$$F_{k+1}(A) \leq F_1(A)^{1/(k+1)} F_k(A)^{k/(k+1)},$$

which is (25).

We now have to show that for any two integers q, p ($q > p \geq 1$) related by $q = kp$ (k an integer) $F_q(A) \leq F_p(A)$. It is sufficient to prove

$$(26) \quad F_k(A) \leq F_1(A) \quad (k = 1, 2, \dots).$$

For since $F_{kp}^p(A) = F_k(A^p)$, (26) implies

$$F_q^p(A) = F_{kp}^p(A) = F_k(A^p) \leq F_1(A^p) = F_p^p(A).$$

(26) is trivial for $k = 1$, therefore using the inequality (25), (26) is seen to be true for $k = 1, 2, \dots$.

The proof for $G_p(A)$ goes in a similar way. For $\nu = 1, 2, \dots$ we have

$$\sum_{\mu=1}^n |a_{\nu\mu}^{(k+1)}| = \sum_{\mu=1}^n \left| \sum_{\lambda=1}^n a_{\nu\lambda}^{(k)} a_{\lambda\mu} \right| \leq \sum_{\lambda=1}^n \left(\sum_{\mu=1}^n |a_{\nu\lambda}^{(k)}| |a_{\lambda\mu}| \right) \leq R_1 R_k.$$

Choosing ν such that $\sum_{\mu=1}^n |a_{\nu\mu}^{(k+1)}| = R_{k+1}$, we get

$$G_{k+1}^{k+1}(A) \leq G_1(A)G_k^k(A),$$

that is,

$$(27) \quad G_{k+1}(A) \leq G_1(A) \left[\frac{G_k(A)}{G_1(A)} \right]^{k/(k+1)} \quad (k = 1, 2, \dots).$$

Again, since $G_k(A^\nu) = G_{kp}^p(A)$ it is sufficient to prove $G_k(A) \leq G_1(A)$ ($k = 1, 2, \dots$) which can be done in the same way as before using (27) instead of (25). This completes the proof of Theorem 5.

In a second note we will extend the results obtained in this section to more general norms.

REFERENCES

1. A. C. AITKEN, *Determinants and matrices*, University Mathematical Texts, vol. 1(1939).
2. ALFRED BRAUER, *Limits for the characteristic roots of a matrix*, this Journal, vol. 13(1946), pp. 387-395.
3. A. B. FARNELL, *Limits for the field of values of a matrix*, American Mathematical Monthly, vol. 52(1945), pp. 488-493.
4. R. A. FRAZER, W. J. DUNCAN, A. R. COLLAR, *Elementary matrices*, Cambridge, 1938.
5. G. H. HARDY, J. E. LITTLEWOOD, G. PÓLYA, *Inequalities*, Cambridge, 1934.
6. RUFUS OLDENBURGER, *Infinite powers of matrices and characteristic roots*, this Journal, vol. 6(1940), pp. 357-361.
7. I. SCHUR, *Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen*, Mathematische Annalen, vol. 66(1909), pp. 488-510.
8. OLGA TAUSKY, *Analytical methods in hypercomplex systems*, Compositio Mathematica, vol. 3(1936), pp. 399-407.
9. OLGA TAUSKY AND JOHN TODD, *Infinite powers of matrices*, Journal of the London Mathematical Society, vol. 17(1942), pp. 146-151.
10. H. W. TURNBULL, A. C. AITKEN, *An introduction to the theory of canonical matrices*, second edition, London and Glasgow, 1945.
11. J. H. M. WEDDERBURN, *The absolute value of the product of two matrices*, Bulletin American Mathematical Society, vol. 31(1925), pp. 304-308.
12. J. H. M. WEDDERBURN, *Lectures on matrices*, American Mathematical Society Colloquium Publications, vol. 17(1934).

UNIVERSITY OF BASLE,
SWITZERLAND.



THE ASYMPTOTIC BEHAVIOUR OF POWERS OF MATRICES. II.

BY WERNER GAUTSCHI

This note is an addendum to our paper [3]. We will extend the results obtained in II of [3] by introducing more general norms, and from this we derive further sequences converging towards ω_A . The enumeration of the equations, theorems and sections will be continued from [3].

III. Generalizations

1. To any column-vector $x = (x_1, \dots, x_n)$ of the n -dimensional complex Euclidean space let a number $\phi(x)$, called the *norm* of x , be assigned, satisfying the following three conditions:

- (i) $\phi(x) > 0$ except for the null vector $x = 0$, for which $\phi(0) = 0$,
- (ii) $\phi(\lambda x) = |\lambda| \phi(x)$ for any complex scalar λ ,
- (iii) $\phi(x + y) \leq \phi(x) + \phi(y)$.

Furthermore, suppose that $\phi(x)$ is bounded over the set of vectors with Euclidean length $|x| = 1$,

$$(iv) \quad \phi(x) \leq C \quad (|x| = 1).$$

Let a function of the vector x , $\psi(x)$, satisfy (i) and (ii) and be bounded from below by a *positive* constant for all vectors x with $|x| = 1$,

$$(v) \quad \psi(x) \geq c > 0 \quad (|x| = 1).$$

Then for an $n \times n$ matrix $A = (a_{\nu\mu})$ the ratio $\phi(Ax)/\psi(x)$ remains bounded over the set of all vectors $x \neq 0$; we may therefore define its least upper bound

$$(28) \quad \Omega_{\phi, \psi}(A) \equiv \sup_{x \neq 0} \frac{\phi(Ax)}{\psi(x)}$$

as the (upper) *norm* of A induced by ϕ and ψ . (Compare for this definition A. Ostrowski [5].) $\Omega_{\phi, \psi}(A)$ is a special case of the most general norm $\Omega(A)$ defined by the three properties:

- (vi) $\Omega(A) > 0$ except when $A = 0$, in which case $\Omega(0) = 0$,
- (vii) $\Omega(\lambda A) = |\lambda| \Omega(A)$ for any complex scalar λ ,
- (viii) $\Omega(A + B) \leq \Omega(A) + \Omega(B)$, A, B being $n \times n$ matrices.

If in particular we take $\psi(x) = \phi(x)$ assuming of course that ϕ satisfies (v), $\Omega_{\phi} \equiv \Omega_{\phi, \phi}$ also satisfies

$$(ix) \quad \Omega_{\phi}(AB) \leq \Omega_{\phi}(A)\Omega_{\phi}(B) \quad (\Omega_{\phi} \equiv \Omega_{\phi, \phi}),$$

Received November 5, 1952. This note is part of the author's doctoral dissertation presented to the University of Basle, Switzerland.

since

$$\frac{\phi(ABx)}{\phi(x)} = \frac{\phi(A(Bx))}{\phi(Bx)} \frac{\phi(Bx)}{\phi(x)}.$$

We mention the following three examples, where $\Omega_{\phi, \psi}$ can be given explicitly in terms of characteristic values of A (compare *e.g.* [1; Chapter 1], [2] or [6]):

$$(a) \quad \phi(x) = \psi(x) = \max_{\nu=1, \dots, n} |x_{\nu}|, \quad \Omega_{\phi, \psi} = \max_{\nu=1, \dots, n} \sum_{\mu=1}^n |a_{\nu\mu}|.$$

$$(b) \quad \phi(x) = \psi(x) = \sum_{\nu=1}^n |x_{\nu}|, \quad \Omega_{\phi, \psi} = \max_{\mu=1, \dots, n} \sum_{\nu=1}^n |a_{\nu\mu}|.$$

$$(c) \quad \phi(x) = \psi(x) = |x| = \left(\sum_{\nu=1}^n |x_{\nu}|^2 \right)^{\frac{1}{2}}, \quad \Omega_{\phi, \psi} = \rho_{\max},$$

where $\rho_{\max}^2$ denotes the dominant eigenvalue of the matrix A^*A . Since $\inf_{x \neq 0} |Ax|/|x|$ is equal to $\rho_{\min}$, $\rho_{\min}^2$ being the smallest eigenvalue of A^*A , we can estimate Ω_{ϕ} in terms of $\rho_{\max}$, $\rho_{\min}$:

$$(29) \quad \frac{c}{C} \rho_{\min} \leq \Omega_{\phi}(A) \leq \frac{C}{c} \rho_{\max}.$$

Indeed, from (iv) and (v) it follows that

$$\frac{c}{C} \rho_{\min} \leq \frac{c}{C} \frac{|Ax|}{|x|} \leq \frac{\phi(Ax)}{\phi(x)} \leq \frac{C}{c} \frac{|Ax|}{|x|} \leq \frac{C}{c} \rho_{\max}.$$

Another class of examples is given by

$$(d) \quad \phi(x) = (\bar{x}' H x)^{\frac{1}{2}}, \quad \psi(x) = (\bar{x}' K x)^{\frac{1}{2}},$$

where H, K denote two positive definite Hermitian matrices. A detailed discussion of the norm $\Omega_{H, K}(A)$ induced by this choice of ϕ, ψ will be given elsewhere [4].

$$(e) \quad \phi(x) = \left(\sum_{\nu=1}^n |x_{\nu}|^r \right)^{1/r}, \quad \psi(x) = \left(\sum_{\nu=1}^n |x_{\nu}|^{r'} \right)^{1/r'}, \quad \frac{1}{r} + \frac{1}{r'} = 1 \quad (1 \leq r \leq \infty).$$

We will denote the corresponding norm of A by $\Omega_{r, r'}(A)$. This norm can easily be estimated from above as follows: put $y = Ax$, $y = (y_1, \dots, y_n)$; then by Hölder's inequality

$$|y_{\nu}|^r \leq \left(\sum_{\mu=1}^n |a_{\nu\mu}|^r \right) \left(\sum_{\mu=1}^n |x_{\mu}|^{r'} \right)^{r/r'} \quad (\nu = 1, \dots, n).$$

Hence summing over $\nu = 1, \dots, n$ and taking the r -th root on both sides we get

$$(30) \quad \phi(y) \leq \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}|^r \right)^{1/r} \psi(x),$$

that is,

$$(31) \quad \Omega_{r, r'}(A) \leq \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}|^r \right)^{1/r}.$$

The limiting cases $r = \infty$ or $r' = \infty$ of (31) follow from (30) by letting $r \rightarrow \infty$, $r' \rightarrow \infty$ respectively in (30) (compare [5], in particular p. 788).

2. The arguments by which we proved the right-hand side of (6) remain unchanged if we use instead of $N(A)$ any norm $\Omega(A)$ satisfying (vi)-(viii). Indeed, the steps leading from (10) to (11) and (11a) can still be applied with Ω instead of N , since they only require (vi)-(viii). We thus obtain the following extension of (6):

THEOREM 7. *Let Ω be a norm for $n \times n$ matrices with the properties (vi)-(viii). Then under the assumptions and with the notations of Theorem 1 we have*

$$(32) \quad \Omega(A^p) \leq cp^{k-1} \left(\sum_{\nu=1}^n |\lambda_{\nu}|^{2p} \right)^{\frac{1}{2}} \quad (p = 1, 2, \dots),$$

where c is a certain positive constant depending only on A and Ω .

Obviously the left-hand side of (6) with Ω instead of N is not true in general, since (vi)-(viii) still hold if Ω is replaced by $\epsilon\Omega$ for any $\epsilon > 0$.

3. In analogy to (22) we put

$$(33) \quad \Phi_p(A) \equiv \Omega_{\phi, \psi}^{1/p}(A^p) \quad (p = 1, 2, \dots),$$

where, for the sake of simplicity, we have dropped the indices ϕ, ψ on the left-hand side. Then from Theorem 7 we deduce the following theorem which contains Theorem 4 and the corollary of Theorem 3 as special cases:

THEOREM 8. *For any ϕ, ψ as defined in 1 we have*

$$(34) \quad \Phi_p(A) \rightarrow \omega_A \quad (p \rightarrow \infty).$$

Moreover, if $\phi(x) = \psi(x)$ then all $\Phi_p(A)$ are bounded from below by ω_A :

$$(35) \quad \Phi_p(A) \geq \omega_A \quad (p = 1, 2, \dots).$$

Proof. Denote by λ an eigenvalue of A with $|\lambda| = \omega_A$ and by x an eigenvector corresponding to λ . Then obviously $\lambda^p x = A^p x$ ($p = 1, 2, \dots$) and from (ii) and (28) it follows that

$$(36) \quad \begin{aligned} \omega_A^p \phi(x) &= \phi(A^p x) \leq \Omega_{\phi, \psi}(A^p) \psi(x), \\ \omega_A \left(\frac{\phi(x)}{\psi(x)} \right)^{1/p} &\leq \Phi_p(A) \end{aligned} \quad (p = 1, 2, \dots).$$

Since (vi)-(viii) hold for $\Omega = \Omega_{\phi, \psi}$ we can combine (36) with (32) and get

$$\omega_A \left(\frac{\phi(x)}{\psi(x)} \right)^{1/p} \leq \Phi_p(A) \leq (cp^{k-1})^{1/p} \left(\sum_{\nu=1}^n |\lambda_{\nu}|^{2p} \right)^{1/2p}$$

whence (34) follows by making $p \rightarrow \infty$.

If $\phi(x) = \psi(x)$, then (35) follows directly from (36).

Again, from Theorem 8 it follows that $\Phi_p(A)$ cannot be considered as a norm for matrices, since the triangular inequality is not satisfied in general, the reason being the same as that given in [3; II].

It is interesting to notice that Theorem 5 can also be extended in the same way:

THEOREM 9. *If $p_\nu > 0$ ($\nu = 1, 2, \dots$) is a strictly increasing sequence of integers such that every p_ν is divisible by $p_{\nu-1}$ ($\nu = 2, 3, \dots$), and if we take $\phi(x) = \psi(x)$, then the sequence $\Phi_{p_\nu}(A)$ as defined by (33) is monotonically decreasing towards ω_A as $\nu \rightarrow \infty$.*

Proof. The proof is essentially the same as for Theorem 5. Since $\phi(x) = \psi(x)$ we can apply (ix) to $A^{k+1} = A \cdot A^k$, which gives

$$\Omega_\phi(A^{k+1}) \leq \Omega_\phi(A) \Omega_\phi(A^k),$$

whence

$$(37) \quad \begin{aligned} \Phi_{k+1}^{k+1}(A) &\leq \Phi_1(A) \Phi_k^k(A), \\ \Phi_{k+1}(A) &\leq \Phi_1(A) \left(\frac{\Phi_k(A)}{\Phi_1(A)} \right)^{k/k+1} \quad (k = 1, 2, \dots). \end{aligned}$$

Again we have to show that for any two integers q, p ($q > p \geq 1$) related by $q = kp$ (k an integer) $\Phi_q(A) \leq \Phi_p(A)$. Since $\Phi_k(A^p) = \Phi_{kp}^p(A)$, it is sufficient to prove

$$\Phi_k(A) \leq \Phi_1(A) \quad (k = 1, 2, \dots),$$

which follows from (37) by the induction argument.

4. By means of Theorems 7 and 8 we now extend Theorem 3. For any real number $r \geq 1$ let the norm $N_r(A)$ be defined by

$$(38) \quad N_r(A) = \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}|^r \right)^{1/r} \quad (1 \leq r \leq \infty).$$

For $\Omega = N_r$ obviously (vi) and (vii) are satisfied while (viii) follows from Minkowski's inequality. Hence Theorem 7 is applicable and by (31), applied to the matrix A^p , we get

$$(39) \quad \Omega_{r, r'}(A^p) \leq N_r(A^p) \leq cp^{k-1} \left(\sum_{\nu=1}^n |\lambda_\nu|^{2p} \right)^{\frac{1}{2}} \quad (p = 1, 2, \dots).$$

On the other hand $\Omega_{r, r'}^{1/p}(A^p) \rightarrow \omega_A$ ($p \rightarrow \infty$) by Theorem 8. Thus by taking the p -th root on both sides of (39) and letting $p \rightarrow \infty$ the following theorem is obtained:

THEOREM 10. *For any $r \geq 1$ we have*

$$N_r^{r^{1/p}}(A^p) \rightarrow \omega_A \quad (p \rightarrow \infty).$$

The limiting cases $r = 1, \infty$ of Theorem 10 are of particular interest, so that we may state them explicitly as

COROLLARY. Let A be an $n \times n$ (real or complex) matrix and put $A^p = (a_{\nu\mu}^{(p)})$,

$$L_p = \left(\sum_{\nu, \mu=1}^n |a_{\nu\mu}^{(p)}| \right)^{1/p}, \quad M_p = \left(\max_{\nu, \mu=1, \dots, n} |a_{\nu\mu}^{(p)}| \right)^{1/p} \quad (p = 1, 2, \dots).$$

Then the sequences L_p and M_p tend to ω_A as $p \rightarrow \infty$.

REFERENCES

1. V. N. FADDEEVA, *Computational methods of linear algebra* (Russian), Moscow-Leningrad, 1950.
2. V. N. FADDEEVA, *Basic material from linear algebra* (Translation of Chapter 1 of [1]), National Bureau of Standards, Washington, D. C.—Report 1644, 1952.
3. WERNER GAUTSCHI, *The asymptotic behaviour of powers of matrices*, this Journal, vol. 20(1953), pp. 127–140.
4. WERNER GAUTSCHI, *Bounds of matrices with regard to an Hermitian metric*, to appear in *Compositio Mathematica*.
5. ALEXANDRE OSTROWSKI, *Un nouveau théorème d'existence pour les systèmes d'équations*, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (Paris)*, vol. 232(1951), pp. 786–788.
6. ALEXANDRE OSTROWSKI, *Lectures on solutions of equations and systems of equations*, Working paper CL-52-2, Washington, D. C., 1952.

UNIVERSITY OF BASLE,
SWITZERLAND.



Bounds of matrices with regard to an Hermitian metric ¹⁾

by

Werner Gautschi.

§ 1. The bounds $\Omega_{H,K}$, $\omega_{H,K}$.

1. *Introduction.* In various questions concerning the solutions of systems of equations and the errors made by rounding off, the following definition of upper and lower bounds $\Omega(A)$, $\omega(A)$ of a matrix A has frequently been used:

$$\Omega(A) = \underset{\nu}{\text{Max}} \sqrt{\varrho_{\nu}}, \quad \omega(A) = \underset{\nu}{\text{Min}} \sqrt{\varrho_{\nu}}, \quad (1)$$

where ϱ_{ν} denote the eigenvalues of $\bar{A}'A$ (cf. e.g. [8] p. 1042 ff., or [9], p. 787, for the special case in which φ, χ are Euclidean lengths). In this paper we will discuss a generalization of this definition introducing as "parameters" two positive definite Hermitian matrices H, K . If H, K vary independently, the generalized bounds $\Omega_{H,K}(A)$, $\omega_{H,K}(A)$ can in general take values in the whole range $(0, \infty)$ (cf. § 1, section 3(vi)); to obtain appropriate values one has to couple H, K in some way. This can be done very naturally when A is an $n \times n$ matrix, by taking $K = H$. The bounds $\Omega_{H,H} \equiv \Omega_H$, $\omega_{H,H} \equiv \omega_H$ are in fact often more favourable for A than (1), but at the same time their actual calculation is considerably more difficult, as is shown by the examples given in § 2. If, however, A contains only a few non-vanishing elements, $\Omega_{H,K}(A)$ can fairly well be estimated from above by means of our theorem 2 in § 3, section 1, which generalizes a theorem due to W. Ledermann [7]. We will also make use of the theorem 2 in § 3, section 2, to determine both $\underset{H}{\text{Inf}} \Omega_H(A)$ and $\underset{H}{\text{Sup}} \omega_H(A)$, where H runs through all positive definite matrices. In sections 2 and 3 of § 1 we give the exact definitions and a few elementary properties of $\Omega_{H,K}(A)$ and $\omega_{H,K}(A)$, while in section 4 of § 1 a property not quite so trivial is proved.

¹⁾ This paper is part of the thesis for the Dr. phil.-degree at the University of Basle, Switzerland.

The idea of relating lengths of vectors to a positive definite Hermitian matrix H has recently been applied to the solution of linear equations $Ax = b$ by M. R. Hestenes and M. L. Stein [6]. Their main problem is to minimize the " H -length" of the residual vectors $r(x) = b - Ax$. Our definition of $\Omega_{H,K}(A)$, $\omega_{H,K}(A)$ involves a similar extremum problem, but (in contrast to [6]) with a *side condition*.

In defining the H -length of a vector we make use of the "scalar product" (x, y) with regard to H of two vectors x, y , as given e.g. in H. L. Hamburger and M. E. Grimshaw [4], p. 153. Such products (x, y) have also recently been used by W. Givens [2] to obtain theorems on the fields of values of a square matrix, which considerably extend the well known results due to O. Toeplitz [12] and F. Hausdorff [5].

I am very much indebted to Prof. Dr. A. Ostrowski for having most kindly allowed me to see through the manuscript of the yet unpublished book [10] from which I received many suggestions. In particular a chapter of [10] on the bounds (1) was the starting point of our investigations, which rather closely follow the disposition of this chapter.

2. Notations and definitions. Let $A = (a_{\mu\nu})$ ($\mu = 1, \dots, m$; $\nu = 1, \dots, n$) be an $m \times n$ matrix with real or complex elements $a_{\mu\nu}$. By $A^* = \bar{A}'$ we denote its conjugate-transpose and by $A^{(p)}$ ($p = 1, \dots, \text{Min}(m, n)$) its p^{th} compound matrix, i.e. the $\binom{m}{p} \times \binom{n}{p}$ matrix consisting of all minors of A of order p . The groups of p rows and columns which form the minors are supposed to be arranged in lexicographical order. We have to use the following rules concerning $A^{(p)}$:

$$(AB)^{(p)} = A^{(p)}B^{(p)}, \quad (A^*)^{(p)} = (A^{(p)})^*, \quad (2)$$

if the product AB exists (cf. e.g. [1], p. 90ff). The first relation in (2) (the so called *Binet-Cauchy* theorem) is readily extended to more than two factors. Further if $m = n$ and A^{-1} exists, from (2) (with $B = A^{-1}$) it follows that

$$(A^{-1})^{(p)} = (A^{(p)})^{-1}. \quad (3)$$

$\text{tr } A$ will denote the trace $\sum_{\nu} a_{\nu\nu}$ of a square matrix A , λ_A an eigenvalue of A and $|\lambda_A|^{\max}$, $|\lambda_A|^{\min}$ respectively the maximal, minimal modulus of the eigenvalues of A .

By x, y etc. we denote *column*-vectors of a k -dimensional complex Euclidean space, by x^* the conjugate-transposed row-vector

$\tilde{x}'$ and by $|x|$ the Euclidean length of x . In order to introduce an Hermitian metric we define the scalar product (x, y) of two vectors x, y by

$$(x, y) \equiv y^* H x \quad (H > 0), \quad (4)$$

where H is an Hermitian matrix of order k ; the meaning of the relation $H > 0$ is that H is positive definite. In particular (x, x) is real and ≥ 0 with $(x, x) = 0$ only when $x = 0$. We therefore define

$$||x|| \equiv \sqrt{(x, x)} \quad (5)$$

as the *norm of x with regard to H* . Sometimes we add the subscript H and write $||x||_H$ instead of $||x||$. By routine arguments (cf. e.g. [11], p. 5, [3], p. 90—92, or [4], p. 4—5) the following three properties of $||x||$ are obtained:

$$\left. \begin{aligned} ||x|| &\geq 0 \text{ with equality if and only if } x = 0 \\ ||\gamma x|| &= |\gamma| ||x|| \quad (\gamma \text{ any complex scalar}) \\ ||x + y|| &\leq ||x|| + ||y||. \end{aligned} \right\} \quad (6)$$

Now let A be an $m \times n$ matrix and $H > 0, K > 0$ be Hermitian matrices of orders m, n respectively; we then define the upper and lower bounds $\Omega_{H,K}(A), \omega_{H,K}(A)$ of A by

$$\left. \begin{aligned} \Omega_{H,K}(A) &= \text{Max}_{||x||_K=1} ||Ax||_H = \left(\text{Max}_{||x||_K=1} x^* A^* H A x \right)^{\frac{1}{2}}, \\ \omega_{H,K}(A) &= \text{Min}_{||x||_K=1} ||Ax||_H = \left(\text{Min}_{||x||_K=1} x^* A^* H A x \right)^{\frac{1}{2}}. \end{aligned} \right\} \quad (7)$$

If in particular $m = n$ and $H = K$ we write $\Omega_{H,H} \equiv \Omega_H, \omega_{H,H} \equiv \omega_H$. The definition (7) can also (partly) be expressed in terms of Euclidean lengths: Let K be transformed to a diagonal matrix by the unitary matrix U :

$$D = U^* K U = \text{Diag}(k_1, \dots, k_n), \quad U^* U = I_n, \quad (8)$$

where I_n is the $n \times n$ unity matrix. Since $k_\nu > 0$ ($\nu = 1, \dots, n$) D can further be reduced to I_n by multiplying on the right and left by $\Delta = \text{Diag}\left(\frac{1}{\sqrt{k_1}}, \dots, \frac{1}{\sqrt{k_n}}\right)$:

$$\Delta U^* K U \Delta = I_n, \quad \Delta = \text{Diag}\left(\frac{1}{\sqrt{k_1}}, \dots, \frac{1}{\sqrt{k_n}}\right). \quad (9)$$

If we now apply to x the substitution $x = U \Delta y$ we get

$$||Ax||_H^2 = x^* A^* H A x = y^* \Delta U^* A^* H A U \Delta y \quad (x = U \Delta y)$$

and by (9)

$$x^*Kx = y^*\Delta U^*KU\Delta y = y^*y.$$

Hence $\|x\|_K = 1$ implies $|y| = 1$ and viceversa; we therefore have

$$\mathfrak{F} \left\| Ax \right\|_H^2 = \mathfrak{F} y^*By, \quad B = \Delta U^*A^*HAU\Delta, \quad (10)$$

$$\|x\|_K=1 \quad |y|=1$$

where $\mathfrak{F}$, $\mathfrak{F}$ denote the fields of values over the sets of vectors $\|x\|_K=1$, $|y|=1$

x, y with $\|x\|_K = 1$, $|y| = 1$ respectively. Since B is non-negative definite, from (10) we see that both Max, Min in (7) actually exist and

$$\Omega_{H,K}^2(A) = \lambda_B^{\max}, \quad \omega_{H,K}^2(A) = \lambda_B^{\min}, \quad B = \Delta U^*A^*HAU\Delta. \quad (11)$$

If in particular we take $H = I_m$, $K = I_n$, so that clearly $U = \Delta = I_n$, we obtain the bounds defined in (1).

Throughout this paper we denote respectively by $h_1, \dots, h_m > 0$, $k_1, \dots, k_n > 0$ the eigenvalues (not necessarily distinct and arranged in any order) of H, K and we put $h' = \text{Max}_{\mu=1, \dots, m} h_\mu$, $h'' = \text{Min}_{\mu=1, \dots, m} h_\mu$; $k' = \text{Max}_{\nu=1, \dots, n} k_\nu$, $k'' = \text{Min}_{\nu=1, \dots, n} k_\nu$.

3. *Elementary properties of $\Omega_{H,K}$, $\omega_{H,K}$.* If not otherwise stated in this section $A, H > 0, K > 0$ are respectively $m \times n$, $m \times m$, $n \times n$ matrices.

(i) The following properties of $\Omega_{H,K}$, $\omega_{H,K}$ are immediate consequences of (6) and (7):

$$\Omega_{H,K}(\gamma A) = |\gamma| \Omega_{H,K}(A), \quad \omega_{H,K}(\gamma A) = |\gamma| \omega_{H,K}(A) \quad (\gamma \text{ any complex scalar}) \quad (12)$$

$$\Omega_{H,K}(A+B) \leq \Omega_{H,K}(A) + \Omega_{H,K}(B), \quad \omega_{H,K}(A+B) \geq \omega_{H,K}(A) - \Omega_{H,K}(B),$$

$$\Omega_{H,K}(A) = 0 \text{ if and only if } A = 0$$

$$\omega_{H,K}(A) = 0 \text{ if and only if the rank of } A \text{ is } < n. \quad (13)$$

(ii) Obviously we can also write

$$\Omega_{H,K}(A) = \text{Max}_{x \neq 0} \frac{\|Ax\|_H}{\|x\|_K}, \quad \omega_{H,K}(A) = \text{Min}_{x \neq 0} \frac{\|Ax\|_H}{\|x\|_K}, \quad (14)$$

so that for any $m \times n$ matrix C :

$\|Cx\|_H \leq \Omega_{H,K}(C) \|x\|_K$, $\|Cx\|_H \geq \omega_{H,K}(C) \|x\|_K$. Hence, if $A, B, L > 0$ respectively are $m \times l$, $l \times n$, $l \times l$ matrices, we have

$$\Omega_{H,K}(AB) \leq \Omega_{H,L}(A) \Omega_{L,K}(B), \quad \omega_{H,K}(AB) \geq \omega_{H,L}(A) \omega_{L,K}(B). \quad (15)$$

On the other hand, for any vector x with $Bx \neq 0$

$$\frac{\|ABx\|_H}{\|x\|_K} = \frac{\|A(Bx)\|_H}{\|Bx\|_L} \frac{\|Bx\|_L}{\|x\|_K} \quad (Bx \neq 0). \quad (16)$$

Suppose now that for the vector x : $\frac{\|Bx\|_L}{\|x\|_K} = \Omega_{L,K}(B)$. Then by (16) and (14)

$$\Omega_{H,K}(AB) \geq \frac{\|ABx\|_H}{\|x\|_K} = \frac{\|A(Bx)\|_H}{\|Bx\|_L} \Omega_{L,K}(B) \geq \omega_{H,L}(A) \Omega_{L,K}(B).$$

Similarly, if B is of rank n , from (16) we deduce $\omega_{H,K}(AB) \leq \Omega_{H,L}(A) \omega_{L,K}(B)$. If B is of rank $< n$, then the same holds for AB (cf. e.g. [1], p. 96—97) and therefore $\omega_{H,K}(AB) = \omega_{L,K}(B) = 0$. Thus we can extend (15) as follows:

$$\Omega_{H,K}(AB) \geq \omega_{H,L}(A) \Omega_{L,K}(B), \quad \omega_{H,K}(AB) \leq \Omega_{H,L}(A) \omega_{L,K}(B). \quad (17)$$

(iii) Suppose that $m = n$ and A^{-1} exists; then putting $x = A^{-1}y$ we see that

$$\mathfrak{F}_{x \neq 0} \frac{\|Ax\|_H}{\|x\|_K} = \mathfrak{F}_{y \neq 0} \frac{\|y\|_H}{\|A^{-1}y\|_K} = \mathfrak{F}_{y \neq 0} \left(\frac{\|A^{-1}y\|_K}{\|y\|_H} \right)^{-1} \quad (x = A^{-1}y).$$

Hence in using (14) we get

$$\Omega_{H,K}(A) = \frac{1}{\omega_{K,H}(A^{-1})}, \quad \omega_{H,K}(A) = \frac{1}{\Omega_{K,H}(A^{-1})}. \quad (18)$$

(iv) Let S, T be two nonsingular matrices of orders m, n respectively; then we have

$$\Omega_{H,K}(A) = \Omega_{S^*HS, T^*KT}(S^{-1}AT), \quad \omega_{H,K}(A) = \omega_{S^*HS, T^*KT}(S^{-1}AT). \quad (19)$$

If in particular $m = n$, $\Omega_{H,K}, \omega_{H,K}$ do not change, if a unitary transformation S is applied both to A, H and K .

Indeed, putting $x = Ty$ we see that the field of values x^*A^*HAx over the set of vectors x with $x^*Kx = 1$ coincides with the field of values

$$y^*T^*A^*HATy = y^*(T^*A^*(S^*)^{-1})(S^*HS)(S^{-1}AT)y$$

taken for all vectors y with $y^*T^*KTy = 1$. Hence (19) follows at once from the definition (7).

(v) Suppose that both A, H and K are respectively the "direct sums" of $A_1, \dots, A_s, H_1, \dots, H_s$ and $K_1, \dots, K_s$, i.e. that in an obvious notation

$$A = \text{Diag}(A_1, \dots, A_s), \quad H = \text{Diag}(H_1, \dots, H_s), \quad K = \text{Diag}(K_1, \dots, K_s),$$

where A_σ is an $m_\sigma \times n_\sigma$, $H_\sigma > 0$ an $m_\sigma \times m_\sigma$ and $K_\sigma > 0$ an $n_\sigma \times n_\sigma$ matrix ($\sigma = 1, \dots, s$). Then we have

$$\Omega_{H,K}(A) = \text{Max}_{\sigma=1,\dots,s} \Omega_{H_\sigma, K_\sigma}(A_\sigma), \quad \omega_{H,K}(A) = \text{Min}_{\sigma=1,\dots,s} \omega_{H_\sigma, K_\sigma}(A_\sigma). \quad (20)$$

In fact, let K_σ be transformed to a diagonal matrix by the unitary matrix U_σ ($\sigma = 1, \dots, s$) and put $U = \text{Diag}(U_1, \dots, U_s)$,

so that clearly (8) holds. Put $\Delta = \text{Diag}\left(\frac{1}{\sqrt{k_1}}, \dots, \frac{1}{\sqrt{k_n}}\right) =$

$\text{Diag}(\Delta_1, \dots, \Delta_s)$, Δ_σ being of the same order as K_σ ($\sigma = 1, \dots, s$). Then obviously the matrix B in (10) is the direct sum of $\Delta_\sigma U_\sigma^* A_\sigma^* H_\sigma A_\sigma U_\sigma \Delta_\sigma$ ($\sigma = 1, \dots, s$), whence (20) follows from (11).

(vi) For every $H > 0$, $K > 0$ we have

$$\sqrt{\frac{h''}{k'}} \omega(A) \leq \omega_{H,K}(A) \leq \Omega_{H,K}(A) \leq \sqrt{\frac{h'}{k''}} \Omega(A), \quad (21)$$

where $\omega(A)$, $\Omega(A)$ are the bounds defined in (1).

Indeed, (21) follows from (14) by putting $y = Ax$ in

$$h'' \|y\|^2 \leq \|y\|_H^2 \leq h' \|y\|^2, \quad k'' \|x\|^2 \leq \|x\|_K^2 \leq k' \|x\|^2.$$

(vii) We have for any eigenvalue λ_A of a square matrix A :

$$\omega_H(A) \leq |\lambda_A| \leq \Omega_H(A). \quad (22)$$

In fact, let x be an eigenvector corresponding to λ_A with $\|x\|_H = 1$. Then $Ax = \lambda_A x$, $\|Ax\|_H = |\lambda_A|$, whence (22) follows directly from (7).

4. For the proof of our first theorem we need the following

LEMMA 1. Let S be an $n \times m$ matrix and T an $m \times n$ matrix. Then, if $m < n$, we have

$$(ST)^{(p)} = 0 \quad (p > m). \quad (23)$$

PROOF. Put $S_0 = (SO_1)$, $T_0 = \begin{pmatrix} T \\ O_2 \end{pmatrix}$, where O_1, O_2 are $n \times (n-m)$, $(n-m) \times n$ zero-matrices respectively. Obviously both S_0 and T_0 are $n \times n$ matrices and $ST = S_0 T_0$. Hence by (2) $(ST)^{(p)} = S_0^{(p)} T_0^{(p)}$, and (23) follows from $S_0^{(p)} = T_0^{(p)} = 0$ ($p > m$).

LEMMA 2. Suppose that $D = \text{Diag}(k_1, \dots, k_n)$ ($k_\nu > 0$),

$\Delta = \text{Diag}\left(\frac{1}{\sqrt{k_1}}, \dots, \frac{1}{\sqrt{k_n}}\right)$ and $G = \text{Diag}(h_1, \dots, h_m)$ ($h_\mu > 0$),

$\Gamma = \text{Diag}\left(\frac{1}{\sqrt{h_1}}, \dots, \frac{1}{\sqrt{h_m}}\right)$. Further let $R = (r_{\mu\nu})$ be an $m \times n$

matrix and put $B = \Gamma R D R^* \Gamma$, $C = \Delta^{-1} R^* G^{-1} R \Delta^{-1}$. Then, if $\varphi(\lambda) = |\lambda I_m - B|$, $\psi(\lambda) = |\lambda I_n - C|$ are the characteristic polynomials of B , C , we have

$$\psi(\lambda) = \lambda^{n-m} \varphi(\lambda).$$

PROOF. Without loss of generality we may assume $m \leq n$. Put $B = (b_{\mu\nu})$ ($\mu, \nu = 1, \dots, m$), $C = (c_{\mu\nu})$ ($\mu, \nu = 1, \dots, n$); by direct multiplication we get

$$b_{\mu\nu} = \frac{1}{\sqrt{h_\mu}} \frac{1}{\sqrt{h_\nu}} \sum_{\sigma=1}^n k_\sigma r_{\mu\sigma} \bar{r}_{\nu\sigma}$$

$$c_{\mu\nu} = \sqrt{k_\mu} \sqrt{k_\nu} \sum_{\tau=1}^m \frac{1}{h_\tau} \bar{r}_{\tau\mu} r_{\tau\nu}.$$

Hence

$$\text{tr } B = \sum_{\mu=1}^m b_{\mu\mu} = \sum_{\mu=1}^m \sum_{\sigma=1}^n \frac{k_\sigma}{h_\mu} |r_{\mu\sigma}|^2 = \sum_{\nu=1}^n \sum_{\tau=1}^m \frac{k_\nu}{h_\tau} |r_{\tau\nu}|^2 = \sum_{\nu=1}^n c_{\nu\nu} = \text{tr } C.$$

We now form the p^{th} compound matrices $B^{(p)}$, $C^{(p)}$ of B , C ; from (2), (3) it follows that

$$B^{(p)} = \Gamma^{(p)} R^{(p)} D^{(p)} (R^{(p)})^* \Gamma^{(p)}$$

$$C^{(p)} = (\Delta^{(p)})^{-1} (R^{(p)})^* (G^{(p)})^{-1} R^{(p)} (\Delta^{(p)})^{-1} \quad (p = 1, \dots, m).$$

Evidently $B^{(p)}$, $C^{(p)}$ are built analogously to B , C . Therefore our first conclusion again is applicable and we get

$$\text{tr } B^{(p)} = \text{tr } C^{(p)} \quad (p = 1, \dots, m). \quad (24)$$

If $m < n$ by the lemma 1 with $S = \Delta^{-1} R^*$, $T = G^{-1} R \Delta^{-1}$ we have

$$C^{(p)} = 0 \quad (p > m). \quad (25)$$

Since generally $(-1)^p \text{tr } A^{(p)}$ is the coefficient of λ^{n-p} in the characteristic polynomial $|\lambda I_n - A|$ of an $n \times n$ matrix A , (cf. e.g. [1], p. 88), our assertion now follows immediately from (24) and (25).

THEOREM 1. Let A be an $m \times n$ matrix and $H > 0$, $K > 0$ be respectively of orders m, n ; then we have

$$\Omega_{K, H}(A^*) = \Omega_{H^{-1}, K^{-1}}(A), \quad (26)$$

and, if $m = n$,

$$\omega_{K, H}(A^*) = \omega_{H^{-1}, K^{-1}}(A). \quad (27)$$

PROOF. Let

$$G = V^* H V = \text{Diag } (h_1, \dots, h_m), \quad V^* V = I_m, \quad (28)$$

$$\Gamma V^* H V \Gamma = I_m, \quad \Gamma = \text{Diag } \left(\frac{1}{\sqrt{h_1}}, \dots, \frac{1}{\sqrt{h_m}} \right) \quad (29)$$

be the equations corresponding to (8), (9), applied to the matrix H . Then in using (8), (28) we have

$$D = U^*KU, D^{-1} = U^*K^{-1}U; G = V^*HV, G^{-1} = V^*H^{-1}V, \quad (30)$$

$$K = UDU^*, K^{-1} = UD^{-1}U^*; H = VGV^*, H^{-1} = VG^{-1}V^*. \quad (31)$$

According to (11) and (28)—(30) we have to examine the eigenvalues of

$$B = \Gamma V^*AKA^*V\Gamma, \quad C = \Delta^{-1}U^*A^*H^{-1}AU\Delta^{-1}.$$

By means of (31) we can write

$$B = \Gamma(V^*AU)D(U^*A^*V)\Gamma = \Gamma RDR^*\Gamma$$

$$C = \Delta^{-1}(U^*A^*V)G^{-1}(V^*AU)\Delta^{-1} = \Delta^{-1}R^*G^{-1}R\Delta^{-1},$$

putting $R = V^*AU$. If we now apply the lemma 2, our assertion follows at once.

COROLLARY 1. *For any square matrix A and $H > 0$ we have*

$$\Omega_H(A^*) = \Omega_{H^{-1}}(A), \quad \omega_H(A^*) = \omega_{H^{-1}}(A). \quad (32)$$

COROLLARY 2. *If A is an Hermitian matrix, then for any $H > 0$*

$$\Omega_H(A) = \Omega_{H^{-1}}(A), \quad \omega_H(A) = \omega_{H^{-1}}(A). \quad (33)$$

§ 2. Examples.

For the sake of simplicity in this section we only consider square $n \times n$ matrices A and we take $H = K$. As to the selection of examples we follow very closely the arrangement given by A. Ostrowski in [10].

(i) Let $A = (a_{\mu\nu})$ be a matrix the only non-vanishing element of which is $a_{ik} \equiv a$. Put $H = (h_{\mu\nu})$, $B = (b_{\mu\nu})$, $U = (u_{\mu\nu})$, where U satisfies (8) and B is the matrix defined in (10). By direct multiplication we get

$$b_{\mu\nu} = |a|^2 h_{ii} \bar{u}_{k\mu} u_{kv} \frac{1}{\sqrt{h_\mu}} \frac{1}{\sqrt{h_\nu}}.$$

If by v we denote the row-vector $\left(\frac{u_{k1}}{\sqrt{h_1}}, \dots, \frac{u_{kn}}{\sqrt{h_n}} \right)$, B can be con-

sidered as the product $|a|^2 h_{ii} v^* v$ and is therefore of rank 1. Hence by (11) we have

$$\Omega_H^2(A) = \lambda_B^{\max} = \text{tr } B = |a|^2 h_{ii} \sum_{\nu=1}^n \frac{|u_{kv}|^2}{h_\nu}.$$

On the other hand, by (31), $h_{ii} = \sum_{v=1}^n h_v |u_{iv}|^2$ and so

$$\Omega_H^2(A) = |a|^2 \sum_{v=1}^n h_v |u_{iv}|^2 \sum_{v=1}^n \frac{|u_{kv}|^2}{h_v}. \quad (34)$$

If in particular H is a diagonal matrix, and therefore $U = I_n$, we get

$$\Omega_H(A) = |a| \sqrt{\frac{h_i}{h_k}}, \quad H = \text{Diag}(h_1, \dots, h_n). \quad (35)$$

Let us in this example discuss, to what extent $\Omega_H(A)$ is determined by the eigenvalues of H . Clearly all Hermitian matrices having the *fixed* eigenvalues $h_1, \dots, h_n > 0$ are obtained by letting U in $H = UDU^*$, $D = \text{Diag}(h_1, \dots, h_n)$, run through *all unitary* $n \times n$ matrices. In the case $i \neq k$, from (34) we can derive the following bounds, between which $\Omega_H^2(A)$ varies:

$$\frac{h''}{h'} \leq \frac{1}{|a|^2} \Omega_H^2(A) \leq \frac{h'}{h''},$$

where the upper and lower bounds are attained by taking in (34) for $(u_{i1}, \dots, u_{in})$, $(u_{k1}, \dots, u_{kn})$ suitable unit vectors. Similarly, if $i = k$, from (34) we see that $\frac{1}{|a|^2} \Omega_H^2(A)$ takes values in a certain closed interval, the left-hand end point of which by (22) is equal to 1.

If on the other hand we let H run through all diagonal matrices, (35) shows that in the case $i \neq k$ the range of $\Omega_H(A)$ is the whole interval $(0, \infty)$, while $\Omega_H(A)$ for $i = k$ is always equal to $|a|$.

Evidently in this example $\omega_H(A) = 0$ by (13).

(ii) Let $A = (a_{\mu\nu})$ be a matrix all elements of which are zero except those lying in the i^{th} row, and put $(a_{i1}, \dots, a_{in}) = (a_1, \dots, a_n) = \alpha$. We suppose that $H = \text{Diag}(h_1, \dots, h_n)$, i.e. $U = I_n$. Then for the matrix B in (10) we have $B = AA^*HA$,

$b_{\mu\nu} = h_i \bar{a}_\mu a_\nu \frac{1}{\sqrt{h_\mu}} \frac{1}{\sqrt{h_\nu}}$ and as in our example (i) the rank of B is equal to 1, so that

$$\Omega_H^2(A) = \text{tr } B = h_i \sum_{v=1}^n \frac{|a_v|^2}{h_v} = |a_i|^2 + h_i \sum_{v=1, v \neq i}^n \frac{|a_v|^2}{h_v}. \quad (36)$$

Clearly we have always $\omega_H(A) = 0$, and, by a suitable choice of H , $\Omega_H(A)$ can take values arbitrarily near to $|a_i| = |\lambda_A|^{\max}$.

(iii) If all elements of the matrix A are equal to $a \neq 0$ and if we take $H = \text{Diag}(h_1, \dots, h_n)$, we have $b_{\mu\nu} = \frac{1}{\sqrt{h_\mu}} \frac{1}{\sqrt{h_\nu}} |a|^2 \sum_{\kappa=1}^n h_\kappa$ and therefore by the Cauchy-Schwarz inequality

$$\Omega_H^2(A) = \text{tr } B = |a|^2 \left(\sum_{\nu=1}^n h_\nu \right) \left(\sum_{\nu=1}^n \frac{1}{h_\nu} \right) \geq |a|^2 n^2. \quad (37)$$

The lower bound for $\Omega_H(A)$, $|a|n = |\lambda_A|^{\max}$, is attained for $H = I_n$, while $\Omega_H(A)$ is not bounded at all from above. On the other hand $\omega_H(A) = 0$.

(iv) Let $A = \text{Diag}(a_1, \dots, a_n)$, $H = \text{Diag}(h_1, \dots, h_n)$. Then $B = \text{Diag}(|a_1|^2, \dots, |a_n|^2)$, so that

$$\Omega_H(A) = \text{Max}_\nu |a_\nu| = |\lambda_A|^{\max}, \quad \omega_H(A) = \text{Min}_\nu |a_\nu| = |\lambda_A|^{\min}. \quad (38)$$

(v) Let $A = (a_{\mu\nu})$ be a matrix all elements of which are zero except those lying in the i^{th} row and k^{th} column, while we have also $a_{ik} = 0$. We further assume H to be a diagonal matrix. In applying to both A and H the same permutation to the rows and columns, whereby in virtue of § 1, section 3(iv), $\Omega_H(A)$, $\omega_H(A)$ are not changed, we can make $k = 1$. Having carried through this transformation we denote by $\alpha = (a_2, \dots, a_n)$ the i^{th} $(n-1)$ -dimensional row-vector of A (without its first element), by $\beta = (b_1, \dots, b_n)$ the first (n) -dimensional column-vector of A (where $b_i = 0$) and we put $H = \text{Diag}(h_1, \dots, h_n)$, $\Delta_1 = \text{Diag}\left(\frac{1}{\sqrt{h_2}}, \dots, \frac{1}{\sqrt{h_n}}\right)$. For the matrix B of (10) we then obtain by direct multiplication (observing that $U = I_n$)

$$B = \begin{pmatrix} \frac{1}{h_1} \beta^* H \beta & 0 \\ 0 & h_i \Delta_1 \alpha^* \alpha \Delta_1 \end{pmatrix}.$$

Since again the $(n-1) \times (n-1)$ matrix in the lower right-hand corner of B is of rank 1, it follows from (11) that

$$\Omega_H(A) = \text{Max}(\Omega_1, \Omega_2), \text{ where } \begin{cases} \Omega_1 = \left(\frac{1}{h_1} \sum_{\substack{\nu=1 \\ \nu \neq i}}^n h_\nu |b_\nu|^2 \right)^{\frac{1}{2}} \\ \Omega_2 = \left(h_i \sum_{\nu=2}^n \frac{1}{h_\nu} |a_\nu|^2 \right)^{\frac{1}{2}} \end{cases}. \quad (39)$$

$$\omega_H(A) = 0 \quad (n > 2).$$

§ 3. A generalization of Ledermann's theorem and the determination of $\inf_{H>0} \Omega_H, \sup_{H>0} \omega_H$.

1. The reason we succeeded to calculate directly Ω_H, ω_H in the examples given in § 2 was that the matrices A contained a sufficiently large number of zeros. We now prove a general theorem which in similar cases always yields an upper bound for $\Omega_{H,K}(A)$ and which is a generalization of a theorem due to W. Ledermann [7]. More precisely:

THEOREM 2. *Let $A = (a_{\mu\nu})$ be an $m \times n$ matrix and denote by α_μ its μ^{th} row-vector; then, if $H = \text{Diag}(h_1, \dots, h_m)$ ($h_\mu > 0$), $K = \text{Diag}(k_1, \dots, k_n)$ ($k_\nu > 0$) and if every column-vector of A contains at most s non-vanishing elements, we have*

$$\Omega_{H,K}^2(A) \leq \sum_{\sigma=1}^s h_{\mu_\sigma} \|\alpha_{\mu_\sigma}\|_{K^{-1}}^2, \quad (40)$$

where the sum on the right-hand side has to be taken over the s largest numbers $h_{\mu_\sigma} \|\alpha_{\mu_\sigma}\|_{K^{-1}}^2$ ($\sigma = 1, \dots, s$) among $h_\mu \|\alpha_\mu\|_{K^{-1}}^2$ ($\mu = 1, \dots, m$).

PROOF. Our proof is essentially the same as that given for the case $H = I_m, K = I_n$ by A. Ostrowski in [10].

Without loss of generality we may assume that

$$h_1 \|\alpha_1\|_{K^{-1}}^2 \geq h_2 \|\alpha_2\|_{K^{-1}}^2 \geq \dots \geq h_m \|\alpha_m\|_{K^{-1}}^2. \quad (41)$$

Indeed, let a permutation P be applied to the rows of A ; if we further permute the rows and the columns of H according to P , by (19) (with $T = I_n$) $\Omega_{H,K}(A)$ does not change and the numbers $h_\mu \|\alpha_\mu\|_{K^{-1}}^2$ ($\mu = 1, \dots, m$) are arranged as required.

Let $\text{Max}_{\|x\|_{K^{-1}}} \|Ax\|_H^2$ be attained for the vector $x = (x_1, \dots, x_n)$

and put $y = Ax, y = (y_1, \dots, y_m)$. For every μ ($\mu = 1, \dots, m$) replace the coordinates x_ν of x for which $a_{\mu\nu} = 0$ by zeros and denote the vector so obtained by $x^{(\mu)}$. Then we have

$$\Omega_{H,K}^2(A) = \|y\|_H^2 = \sum_{\mu=1}^m h_\mu |y_\mu|^2 = \sum_{\mu=1}^m h_\mu |\alpha_\mu x^{(\mu)}|^2. \quad (42)$$

We further put $x^{(\mu)} = (x_1^{(\mu)}, \dots, x_n^{(\mu)})$ ($\mu = 1, \dots, m$). Since by the Cauchy-Schwarz inequality

$$\begin{aligned} |\alpha_\mu x^{(\mu)}|^2 &= \left| \sum_{\nu=1}^n a_{\mu\nu} x_\nu^{(\mu)} \right|^2 \leq \left(\sum_{\nu=1}^n \frac{1}{\sqrt{k_\nu}} |a_{\mu\nu}| \sqrt{k_\nu} |x_\nu^{(\mu)}| \right)^2 \leq \\ &\leq \left(\sum_{\nu=1}^n \frac{1}{k_\nu} |a_{\mu\nu}|^2 \right) \left(\sum_{\nu=1}^n k_\nu |x_\nu^{(\mu)}|^2 \right) = \|\alpha_\mu\|_{K^{-1}}^2 \sum_{\nu=1}^n k_\nu |x_\nu^{(\mu)}|^2, \end{aligned}$$

from (42) we get

$$\begin{aligned}\Omega_{H,K}^2(A) &\leq \sum_{\mu=1}^m h_{\mu} \|\alpha_{\mu}\|_{K^{-1}}^2 \sum_{\nu=1}^n k_{\nu} |x_{\nu}^{(\mu)}|^2 = \\ &= \sum_{\nu=1}^n k_{\nu} \left[\sum_{\mu=1}^m |x_{\nu}^{(\mu)}|^2 h_{\mu} \|\alpha_{\mu}\|_{K^{-1}}^2 \right]. \quad (43)\end{aligned}$$

If $x_{\nu}^{(\mu)} \neq 0$ then

$$x_{\nu}^{(\mu)} = x_{\nu} \quad (x_{\nu}^{(\mu)} \neq 0) \quad (44)$$

and $a_{\mu\nu} \neq 0$. From this and the hypothesis it follows that for any fixed ν at most s of the $x_{\nu}^{(\mu)}$ are $\neq 0$. Therefore taking the sum in brackets on the right-hand side of (43) only over the terms with $x_{\nu}^{(\mu)} \neq 0$ and using (44), (41) we see that

$$\sum_{\mu=1}^m |x_{\nu}^{(\mu)}|^2 h_{\mu} \|\alpha_{\mu}\|_{K^{-1}}^2 \leq |x_{\nu}|^2 \sum_{\sigma=1}^s h_{\sigma} \|\alpha_{\sigma}\|_{K^{-1}}^2 \quad (\nu = 1, \dots, n),$$

whence by (43)

$$\Omega_{H,K}^2(A) \leq \left(\sum_{\nu=1}^n k_{\nu} |x_{\nu}|^2 \right) \left(\sum_{\sigma=1}^s h_{\sigma} \|\alpha_{\sigma}\|_{K^{-1}}^2 \right).$$

This proves our assertion, since $\sum_{\nu=1}^n k_{\nu} |x_{\nu}|^2 = \|x\|_K^2 = 1$.

REMARKS. The theorem of Ledermann is obtained by taking $H = I_m$, $K = I_n$. If in particular we apply (40) with $H = K$ to our examples (i), (ii) and (iv) we obtain respectively as upper bounds $\frac{h_i}{h_k} |a|^2$, $h_i \|\alpha\|_{H^{-1}}^2$, $\max_{\nu} |a_{\nu}|^2$, which all coincide with the corresponding Ω_H^2 .

Even if $s = m$ the theorem 2 is often useful. Take e.g.

$$A = \begin{pmatrix} 2 & 1 & 8 \\ 7 & 0 & 5 \\ 0 & 1 & 3 \end{pmatrix},$$

where the elements of the second column are comparatively small. In order to get favourable bounds for $\Omega_H(A)$ in applying (40), we choose h_2 relatively small. With $H = \text{Diag}(4, 1, 20)$ we obtain $\Omega_H^2(A) \leq 63,3$, while $H = I$ gives $\Omega^2(A) \leq 153$.

2. We now use our theorem 2 to give a refinement of (22):

THEOREM 3. *For any $n \times n$ matrix A we have*

$$\inf_{H>0} \Omega_H(A) = |\lambda_A|^{\max}, \quad \sup_{H>0} \omega_H(A) = |\lambda_A|^{\min}. \quad (45)$$

If in particular A has only simple elementary divisors both $\inf$ and $\sup$ in (45) are attained for suitable matrices $H > 0$.

PROOF. Since for a nonsingular matrix $|\lambda_{A^{-1}}|^{\max} = 1/|\lambda_A|^{\min}$ and by (18) $\omega_H(A) = \frac{1}{\Omega_H(A^{-1})}$, it is sufficient to prove the relation concerning $\text{Inf } \Omega_H$. Let A be transformed to Jordan's canonical form by the nonsingular matrix S :

$$S^{-1}AS = A + C,$$

where A is a diagonal matrix the elements of which are the eigenvalues of A , and C denotes a matrix consisting of zeros except possibly some elements $c_{\mu\nu} = 1$ with $\nu = \mu + 1$. If in (19) we take $T = S$ we have by (22), (12)

$$|\lambda_A|^{\max} \leq \Omega_H(A) = \Omega_K(A + C) \leq \Omega_K(A) + \Omega_K(C), \quad (46)$$

where $K = S^*HS$. It suffices to show that for a suitable choice of K the sum on the right-hand side of (46) is arbitrarily near to $|\lambda_A|^{\max}$. Take $K = \text{Diag}(k_1, \dots, k_n)$; then by (38) $\Omega_K(A) = |\lambda_A|^{\max}$; if on the other hand γ_ν is the ν^{th} row-vector of C ,

then $\|\gamma_\nu\|_{K^{-1}}^2 = \begin{cases} 0 (\gamma_\nu = 0) \\ 1/k_{\nu+1} (\gamma_\nu \neq 0) \end{cases}$. Hence by the theorem 2

$$\Omega_K^2(C) \leq \text{Max}_{\nu=1, \dots, n-1} \frac{k_\nu}{k_{\nu+1}},$$

which obviously can be made as small as we please.

If all elementary divisors of A are simple we have in (46) $C = 0$, $\Omega_K(C) = 0$, so that $\Omega_H(A)$ attains the value $|\lambda_A|^{\max}$ for a suitable matrix $H > 0$.

It is natural to ask whether we could in (45) take Inf , Sup only over the set of all *diagonal* matrices $H > 0$. This is however not true as the following example shows: Take

$$A = \begin{pmatrix} 0 & i & 1 \\ i & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (i^2 = -1),$$

where $|\lambda_A|^{\max} = 0$. If $H = \text{Diag}(h_1, h_2, h_3)$, it follows from § 2, Ex. (v), that

$$\Omega_H(A) = \text{Max}(\Omega_1, \Omega_2), \quad \text{where} \quad \begin{cases} \Omega_1 = \sqrt{\frac{h_2 + h_3}{h_1}} \\ \Omega_2 = \sqrt{h_1 \left(\frac{1}{h_2} + \frac{1}{h_3} \right)} = \sqrt{\frac{h_1(h_2 + h_3)}{h_2 h_3}}. \end{cases}$$

But by the inequality of the arithmetic and geometric mean

$$\Omega_1 \Omega_2 = \frac{h_2 + h_3}{\sqrt{h_2 h_3}} \geq 2,$$

so that certainly $\Omega_H(A) \geq \sqrt{2}$ for all diagonal matrices $H > 0$.

The second statement in theorem 3 can be made more precise by the following

THEOREM 4. *In order that for some matrix $H > 0$*

$$\Omega_H(A) = |\lambda_A|^{\max} \quad (47)$$

it is necessary and sufficient that the elementary divisors corresponding to the eigenvalues of A with maximal modulus are simple. Similarly, if A is a non-singular matrix, we have

$$\omega_H(A) = |\lambda_A|^{\min} \quad (|\lambda_A|^{\min} > 0) \quad (48)$$

for some $H > 0$, if and only if all elementary divisors associated with the eigenvalues of A of minimal modulus are simple.

PROOF. *Necessity:* let λ be an eigenvalue of A of either maximal or minimal modulus having multiple elementary divisors. It then suffices to show that, given a matrix $H > 0$, there always exists a vector x for which

$$\frac{\|Ax\|_H}{\|x\|_H} \begin{cases} < |\lambda_A|^{\min}, & \text{if } |\lambda| = |\lambda_A|^{\min} \\ > |\lambda_A|^{\max}, & \text{if } |\lambda| = |\lambda_A|^{\max}. \end{cases} \quad (49)$$

From Jordan's canonical form of A it is easily seen that under our hypothesis on λ there exist two linearly independent vectors u_1, u_2 such that $Au_1 = \lambda u_1$, $Au_2 = \lambda u_2 + u_1$. Put $v_1 = u_1$, $v_2 = \alpha u_1 + u_2$; in order to make v_1, v_2 orthogonal with respect to H , using the notation (4) we must have

$$(v_2, v_1) = (\alpha u_1 + u_2, u_1) = \alpha(u_1, u_1) + (u_2, u_1) = 0, \\ \alpha = -(u_2, u_1)/(u_1, u_1).$$

Clearly $Av_1 = \lambda v_1$, $Av_2 = \alpha \lambda u_1 + \lambda u_2 + u_1 = \lambda v_2 + v_1$, and so the vectors $w_1 = v_1/\|v_1\|_H$, $w_2 = v_2/\|v_2\|_H$ satisfy

$$Aw_1 = \lambda w_1 \\ Aw_2 = \lambda w_2 + \beta w_1 \quad (\|w_1\|_H = \|w_2\|_H = 1, (w_2, w_1) = 0), \quad (50)$$

where $\beta = \|v_1\|_H/\|v_2\|_H > 0$. We now take

$$x = \gamma w_1 + w_2 \quad (51)$$

and determine the scalar γ in such a way that (49) holds. In fact, by (50)

$$Ax = \gamma \lambda w_1 + \lambda w_2 + \beta w_1 = \lambda x + \beta w_1, \\ \|Ax\|_H^2 = x^* A^* H A x = (\bar{\lambda} x^* + \beta w_1^*) (\lambda H x + \beta H w_1) = \\ = |\lambda|^2 \|x\|_H^2 + 2\Re(\beta \lambda w_1^* H x) + \beta^2.$$

Substituting the expression (51) for x in $w_1^* H x$ we obtain

$$\frac{\|Ax\|_H^2}{\|x\|_H^2} = |\lambda|^2 + \frac{\beta}{\|x\|_H^2} \left\{ 2\Re(\gamma \lambda) + \beta \right\} \quad (\beta > 0).$$

Now (49) certainly holds, if in the case $|\lambda| = |\lambda_A|^{\min}$ we choose γ such that $\Re(\gamma\lambda) < \frac{-\beta}{2}$ and $\gamma = 0$ if $|\lambda| = |\lambda_A|^{\max}$.

Sufficiency: suppose that all eigenvalues with maximal modulus have simple elementary divisors. Let A be transformed to Jordan's canonical form $S^{-1}AS = J = \text{Diag}(J_1, J_2)$, where J_1 is a *diagonal* matrix containing the eigenvalues λ with $|\lambda| = |\lambda_A|^{\max}$. By (19) we have $\Omega_H(A) = \Omega_K(J)$ ($K = S^*HS$). To show that for a suitable matrix $K > 0$: $\Omega_K(J) = |\lambda_A|^{\max}$, take $K = \text{Diag}(K_1, K_2)$, where $K_1, K_2 > 0$ are matrices of the same order as J_1, J_2 respectively and K_1 is a unity matrix. Since $\Omega_{K_1}(J_1) = |\lambda_A|^{\max}$, from (20) we get

$$\Omega_K(J) = \text{Max} \{ |\lambda_A|^{\max}, \Omega_{K_2}(J_2) \}.$$

On the other hand $|\lambda_{J_2}|^{\max} < |\lambda_A|^{\max}$, whence, by theorem 3, K_2 can be chosen such that $\Omega_{K_2}(J_2) < |\lambda_A|^{\max}$.

A similar argument shows that (48) holds for some $H > 0$, if all eigenvalues of A with minimal modulus are simple.

REFERENCES.

A. C. AITKEN.

[1] Determinants and matrices, University Math. Texts, vol. 1, 7th edition. Edinburgh & London 1951.

[2] W. GIVENS.

Fields of values of a matrix, Bull. Amer. Math. Soc., vol. 58 (1952), p. 53.

P. R. HALMOS.

[3] Finite dimensional vector spaces, Annals of Mathematics Studies, nr. 7, Princeton 1942.

(P. 91, line 10: instead of $|y|^2$ read $\|y\|^2$

p. 91, line 15: instead of $R[\alpha\beta(x, y)]$ read $R[\alpha\bar{\beta}(x, y)]$.)

H. L. HAMBURGER and M. E. GRIMSHAW.

[4] Linear Transformations in n -dimensional vector space, Cambridge University Press, Cambridge 1951.

F. HAUSDORFF.

[5] Der Wertevorrat einer Bilinearform, Math. Z., Bd. 3 (1919), pp. 314—316.

M. R. HESTENES and M. L. STEIN.

[6] The Solution of Linear Equations by Minimization, NBS-NAML Report 52—45, Washington, D.C., 1951.

W. LEDERMANN.

[7] On an upper limit for the latent roots of a certain class of matrices, J. Lond. Math. Soc., vol. 12 (1937), pp. 12—18.

J. VON NEUMANN and H. H. GOLDSTINE.

- [8] Numerical inverting of matrices of high order, *Bull. Amer. Math. Soc.*, vol. 53 (1947), pp. 1021—1099.

A. OSTROWSKI.

- [9] Un nouveau théorème d'existence pour les systèmes d'équations, *C. R. Acad. Sci. Paris*, t. 232 (1951), pp. 786—788.
[10] Leçons sur la résolution des systèmes d'équations, to appear in the series of the cahiers scientifiques at Gauthier-Villars, Paris.

M. H. STONE.

- [11] Linear Transformations in Hilbert Space and their applications to Analysis, *Amer. Math. Soc. Coll. Publ.*, vol. XV (1932).

O. TOEPLITZ.

- [12] Das algebraische Analogon zu einem Satze von Fejér, *Math. Z.*, Bd. 2 (1918), pp. 187—197.

(Oblatum 10-2-53).

University of Basle, Switzerland.

CONTENTS

The asymptotic behaviour of powers of matrices, I.

§ 1. The main results

§ 2. Corollaries

The asymptotic behaviour of powers of matrices, II.

§ 3. Generalizations

Bounds of matrices with regard to an Hermitian metric.

§ 1. The bounds $\Omega_{H,K}$, $\omega_{H,K}$

§ 2. Examples

§ 3. A generalization of Ledermann's theorem and the determination of

$\inf \Omega_H$, $\sup \omega_H$

$H > O$ $H > O'$

Curriculum vitae

Am 11. Dezember 1927 wurde ich, *Werner Gautschi*, als Sohn des Kaufmanns Hans Gautschi und seiner Gattin Margaritha, geb. Eugster, in Basel geboren. Mein Heimatort ist Reinach (Kanton Aargau). Im Jahre 1938 trat ich in das Humanistische Gymnasium Basel ein und bestand dort im Frühling 1946 die Maturitätsprüfung. Daraufhin bezog ich die hiesige Universität zum Studium von Mathematik, Astronomie, Physik und Versicherungsmathematik. 1948–1950 war ich Assistent bei Herrn Prof. Ostrowski. Durch seine Vermittlung wurde mir anschliessend, 1950–1951, die Möglichkeit gegeben, während dreier Terms in Cambridge (England) mich weiterzubilden.

In Basel besuchte ich Vorlesungen und Übungen der folgenden Herren Professoren und Privatdozenten: E. Baldinger, E. Batschelet, P. Buchner, M. Fierz, J. O. Fleckenstein, P. Huber, E. Miescher, R. Nevanlinna, A. Ostrowski, E. Von der Pahlen, E. Schubarth, A. Speiser, B. L. van der Waerden, E. Zwinggi.

In Cambridge hörte ich bei Prof. A. S. Besicovitch, Dr. J. C. Burkill, Dr. M. L. Cartwright, Dr. J. W. S. Cassels, Miss M. E. Grimshaw, Mr. P. Hall, Mr. A. E. Ingham, Prof. J. E. Littlewood, Dr. R. A. Rankin, Dr. F. Smithies, Dr. A. J. Ward.

Allen meinen Lehrern bin ich zu grossem Dank verpflichtet. Ganz besonderen Dank schulde ich Herrn Prof. Ostrowski für die mannigfache Förderung, die ich während des Studiums durch ihn erfahren habe. Unter seiner Leitung ist auch die vorliegende Dissertation entstanden, zu deren Ausgestaltung seine wertvollen Ratschläge ganz wesentlich beigetragen haben.

des
ter,
Im
und
hin
tik
war
ung
end

den
et,
R.
ei-

r-
w,
n-
w
r-
n
z

